

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
29 March 2001 (29.03.2001)

PCT

(10) International Publication Number  
WO 01/21795 A2

(51) International Patent Classification<sup>7</sup>: C12N 15/12,  
C07K 14/705, C12N 5/10, 15/62, G01N 33/50, 33/68,  
C12Q 1/68, C07K 16/28, A61K 47/48 // A61P 3/04

(21) International Application Number: PCT/US00/25891

(22) International Filing Date:  
21 September 2000 (21.09.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
09/405,505 23 September 1999 (23.09.1999) US  
09/405,504 23 September 1999 (23.09.1999) US  
09/465,280 16 December 1999 (16.12.1999) US  
09/506,252 17 February 2000 (17.02.2000) US  
09/611,197 6 July 2000 (06.07.2000) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	09/405,505 (CIP)
Filed on	23 September 1999 (23.09.1999)
US	09/405,504 (CIP)
Filed on	23 September 1999 (23.09.1999)
US	09/465,280 (CIP)
Filed on	16 December 1999 (16.12.1999)
US	09/506,252 (CIP)
Filed on	17 February 2000 (17.02.2000)
US	09/611,197 (CIP)
Filed on	6 July 2000 (06.07.2000)

(71) Applicants (for all designated States except US): WHITE-HEAD INSTITUTE FOR BIOMEDICAL RESEARCH [US/US]; Nine Cambridge Center, Cambridge, MA 02142 (US). MILLENNIUM PHARMACEUTICALS, INC. [US/US]; 75 Sidney Street, Cambridge, MA 02139 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): STAHL, Andreas [DE/US]; 12 Farrington Avenue, #2, Allston, MA 02134 (US). HIRSCH, David, J. [US/US]; 17 Rockview Street, #3, Jamaica Plain, MA 02130 (US). LODISH, Harvey, F. [US/US]; 195 Fisher Avenue, Brookline, MA 02446 (US). GIMENO, Ruth, E. [DE/US]; 65 Beverly Road, Wellesley, MA 02481 (US). TARTAGLIA, Louis, A. [US/US]; 32 Manor House Road, Newton, MA 02459 (US).

(74) Agents: HOGLE, Doreen, M. et al.; Hamilton, Brook, Smith & Reynolds, P.C., Two Militia Drive, Lexington, MA 02421 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— Without international search report and to be republished upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: FATTY ACID TRANSPORT PROTEINS

(57) Abstract: A family of fatty acid transport proteins (FATPs) mediate transport of long chain fatty acids (LCFAs) across cell membranes into cells. These proteins exhibit different expression patterns among the organs of mammals. Nucleic acids encoding FATPs of this family, vectors comprising these nucleic acids, as well as the production of FATP proteins in host cells are described. Also described are methods to test FATPs for fatty acid transport function, and methods to identify inhibitors or enhancers of transport function. The altering of LCFA uptake by administering to the mammal an inhibitor or enhancer of FATP transport function of a FATP in the small intestine can decrease or increase calories available as fats, and can decrease or increase circulating fatty acids. The organ specificity of FATP distribution can be exploited in methods to direct drugs, diagnostic indicators and so forth to an organ such as the heart.

WO 01/21795 A2

AP

## FATTY ACID TRANSPORT PROTEINS

## RELATED APPLICATION(S)

This application is a continuation-in-part of U.S. Patent Application Number 09/506,252 filed February 17, 2000 which is a continuation-in-part of U.S. Patent Application Number 09/465,280 filed December 16, 1999 which is a continuation-in-part of U.S. Patent Application Number 09/405,505 filed September 23, 1999, and is a continuation-in-part of U.S. Patent Application Number 09/232,195 filed January 14, 1999, both of which claim the benefit of U.S. Provisional Application Number 60/110,941 filed December 4, 1998; U.S. Provisional Application Number 60/093,491 filed July 20, 1998; and U.S. Provisional Application Number 60/071,374 filed January 15, 1998. This application is also a continuation-in-part of U.S. Patent Application Number 09/405,504 filed September 23, 1999, which is a continuation-in-part of U.S. Patent Application Number 09/232,201 filed January 14, 1999, which claims the benefit of U.S. Provisional Application Number 60/110,941 filed December 4, 1998; U.S. Provisional Application Number 60/093,491 filed July 20, 1998; and U.S. Provisional Application Number 60/071,374 filed January 15, 1998. This application is also a continuation-in-part of U.S. Patent Application Number 09/232,197 filed January 14, 1999, United States Patent Application Number 09/232,200 filed January 14, 1999 and International Application Number PCT/US99/00182 filed January 14, 1999, each of which claims the benefit of U.S. Provisional Application Number 60/110,941 filed December 4, 1998; U.S. Provisional Application Number 60/093,491 filed July 20, 1998; and U.S.



-2-

Provisional Application Number 60/071,374 filed January 15, 1998. The teachings of each of these referenced applications are incorporated herein by reference in their entirety.

#### GOVERNMENT SUPPORT

5                   The invention was supported, in whole or in part, by a grant from the National Heart, Lung, and Blood Institute (HL41484), by National Institutes of Health Grant DK 47618 and National Institutes of Health Grant 5 T32 CA 09541. The United States Government has certain rights in the invention.

#### BACKGROUND OF THE INVENTION

10               Long chain fatty acids (LCFAs) are an important source of energy for most organisms. They also function as blood hormones, regulating key metabolic functions such as hepatic glucose production. Although LCFAs can diffuse through the hydrophobic core of the plasma membrane into cells, this nonspecific transport cannot account for the high affinity and specific transport of LCFAs exhibited by  
15   cells such as cardiac muscle, hepatocytes, enterocytes, and adipocytes. The molecular mechanisms of LCFA transport remains largely unknown. Identifying these mechanisms can lead to pharmaceuticals that modulate fatty acid uptake by the intestine and by other organs, thereby alleviating certain medical conditions (e.g. obesity).

#### 20   SUMMARY OF THE INVENTION

              Described herein is a diverse family of fatty acid transport proteins (FATPs) which are evolutionarily conserved; these FATPs are plasma membrane proteins which mediate transport of LCFAs across the membranes and into cells. Members of the FATP family described herein are present in a wide variety of organisms, from  
25   mycobacteria to humans, and exhibit very different expression patterns in tissues among the organisms. FATP family members are expressed in prokaryotic and eukaryotic organisms and comprise characteristic amino acid domains or sequences which are highly conserved across family members. In addition, the function of the

FATP gene family is conserved throughout evolution, as shown by the fact that the *Caenorhabditis (C). elegans* and mycobacterial FATPs described herein facilitate LCFA uptake when they are overexpressed in COS cells or *Escherichia (E.) coli*, respectively. FATPs are expressed in a wide variety of tissues, including all tissues  
5 which are important to fatty acid metabolism (uptake and processing).

In specific embodiments, FATPs of the present invention are from such diverse organisms as humans (*Homo (H.) sapiens*), mice, (*Mus (M.) musculus*), *F. rubripes*, *C. elegans*, *Drosophila (D.) melanogaster*, *Saccharomyces (S.) cerevisiae*, *Aspergillus nidulans*, *Cochliobolu heterostrophus*, *Magnaporthe grisea* and  
10 *Mycobacterium (M.)*, such as *M. tuberculosis*. As described herein, four novel mouse FATPs, referred to as mmFATP2, mmFATP3, mmFATP4 and mmFATP5, and six human FATPs, referred to as hsFATP1, hsFATP2, hsFATP3, hsFATP4, hsFATP5 and hsFATP6, have been identified. All four novel murine FATPs (mmFATP2-5) and a previously identified murine FATP (renamed herein FATP1)  
15 have orthologs in humans (hsFATP1-5); the sixth human FATP (hsFATP6) does not as yet have a mouse ortholog. The expression patterns of these FATPs vary, as described in detail below.

The present invention relates to FATP family members from prokaryotes and eukaryotes, nucleic acids (DNA, RNA) encoding FATPs, and nucleic acids which  
20 are useful as probes or primers (e.g., for use in hybridization methods, amplification methods) for example, in methods of detecting FATP-encoding genes, producing FATPs, and purifying or isolating FATP-encoding DNA or RNA. Also the subject of this invention are antibodies (polyclonal or monoclonal) which bind an FATP or FATPs; methods of identifying additional FATP family members (for example,  
25 orthologs of those FATPs described herein by amino acid sequence) and variant alleles of known FATP genes; methods of identifying compounds which bind to an FATP, or modulate or alter (enhance or inhibit) FATP function; compounds which modulate or alter FATP function; methods of modulating or altering (enhancing or inhibiting) FATP function and, thus, LCFA uptake into tissues of a mammal (e.g.  
30 human) by administering a compound or molecule (a drug or agent) which increases or reduces FATP activity; and methods of targeting compounds to tissues by

administering a complex of the compound to be targeted to tissues and a component which is bound by an FATP present on cells of the tissues to which the compound is to be targeted. For example, a complex of a drug to be delivered to the liver and a component which is bound by an FATP present on liver cells (e.g., FATP5) can be  
5 administered.

In one embodiment, the present invention relates to modulating or altering (enhancing or inhibiting/reducing) LCFA uptake in the small intestine and, thus, increasing or reducing the number of calories in the form of fats available to an individual. In another embodiment, the present invention relates to inhibiting or  
10 reducing LCFA uptake in the small intestine in order to reduce circulating fatty acid levels; that is, LCFA uptake in the small intestine is reduced and, therefore, circulating (blood) levels are not as high as they otherwise would be. FATP4 has been shown to be expressed in epithelial cells of the small intestine and particularly in the brush border layer of the small intestine. FATP2 has also been shown to be  
15 expressed at low levels in epithelial cells of the small intestine, particularly in the duodenum. In contrast, FATP1, FATP3, FATP5 and FATP6 were not detected in any of the intestinal tissues. Thus, also described herein are FATPs which are present in the epithelial cell layer of the small intestine where they mediate LCFA uptake. These FATPs, particularly FATP4 and also FATP2, are targets for methods  
20 and drugs which block their function or activity and are useful in treating obesity, diabetes and heart disease. The ability of these FATPs to mediate fat uptake can be modulated or altered (enhanced or inhibited), thus modulating fat uptake in the small intestine. This can be done, for example, by administering to an individual, such as a human or other animal, a drug which blocks interaction of LCFAs with FATP4  
25 and/or FATP2 in the small intestine, thus inhibiting LCFA passage into the cells of the small intestine. As a result, fat absorption is reduced and, although the individual has consumed a certain quantity of fat, the LCFAs are not absorbed to the same extent they would have been in the absence of the compound administered.

Thus, one embodiment of this invention is a method of reducing LCFA  
30 uptake (absorption) in the small intestine and, as a result, reducing caloric uptake in the form of fat. A further embodiment is a compound (drug) useful in inhibiting or

reducing fat absorption in the small intestine. In another embodiment, the invention is a method of reducing circulating fatty acid levels by administering to an individual a compound which blocks interactions of LCFAs with FATP4 and/or FATP2 in the small intestine, thus inhibiting LCFA passage into cells of the small intestine. As a  
5 result, fatty acids pass into the circulatory system at a diminished level and/or rate, and circulating fatty acid levels are lower than they would be in the absence of the compound administered. This method is particularly useful for therapy in individuals who are at risk for or have hyperlipidemia. That is, it can be used to prevent the occurrence of elevated levels of lipids in the blood or to treat an  
10 individual in whom blood lipid levels are elevated. Also the subject of this invention is a method of identifying compounds which alter FATP function (and thus, in the case of FATP2 and/or FATP4, alter LCFA uptake in the small intestine).

In another embodiment, the present invention relates to a method of modulating or altering (enhancing or inhibiting) the function of FATP6, which is  
15 expressed at high levels in the heart. A method of inhibiting FATP6 function is useful, for example, in individuals with heart disease, such as ischemia, since reducing LCFA uptake into heart muscle in an individual who has ischemic heart disease, which may be manifested by, for example, angina or heart attack, can reduce symptoms or reduce the extent of damage caused by the ischemia. In this  
20 embodiment, a drug which inhibits FATP6 function is administered to an individual who has had or is having a heart attack, to reduce LCFA uptake by the individual's heart and, as a result, reduce the damage caused by ischemia. In a further embodiment, this invention is a method of targeting a compound, such as a therapeutic drug or an imaging reagent, to heart tissue by administering to an  
25 individual (e.g., a human) a complex of the compound and a component (e.g., a LCFA or LCFA-like compound) which is bound by an FATP (e.g., FATP6) present in cells of heart tissue.

In a further embodiment, LCFA uptake by the liver is modulated or altered (enhanced or reduced), in an individual. For example, a drug which inhibits the  
30 function of an FATP present in liver (e.g., FATP5) is administered to an individual

who is diabetic, in order to reduce LCFA uptake by liver cells and, thus reduce insulin resistance.

The present invention, thus, provides methods which are useful to alter, particularly reduce, LCFA uptake in individuals and, as a result, to alter (particularly  
5 reduce), availability of the LCFAs for further metabolism. In a specific embodiment, the present invention provides methods useful to reduce LCFA uptake and, thus, fatty acid metabolism in individuals, with the result that caloric availability from fats is reduced, and circulating fatty acid levels are lower than they otherwise would be. These methods are useful, for example, as a means of weight  
10 control in individuals, (e.g., humans) and as a means of preventing elevated serum lipid levels or reducing serum lipid levels in humans. FATPs expressed in the small intestine, such as FATP4, are useful targets to be blocked in treating obesity (e.g., chronic obesity) or to be enhanced in treating conditions in which enhanced LCFA uptake is desired (e.g., malabsorption syndrome or other wasting conditions).

15 The identification of this evolutionarily conserved fatty acid transporter family will allow a better understanding of the mechanisms whereby LCFAs traverse the lipid bilayer as well as yield insight into the control of energy homeostasis and its dysregulation in diseases such as diabetes and obesity.

#### BRIEF DESCRIPTION OF THE DRAWINGS

20 The file of this patent contains at least one color photograph. Copies of this patent with color photographs will be provided by the Patent and Trademark Office upon request and payment of necessary fee.

Figure 1 shows the amino acid sequence alignment of FATPs: mmFATP1 (SEQ ID NO:92), mmFATP2 (SEQ ID NO:93), mmFATP3 (SEQ ID NO:94),  
25 mmFATP4 (SEQ ID NO:95), mmFATP5 (SEQ ID NO:96), ceFATPa (SEQ ID NO:97), scFATP (SEQ ID NO:98) and mtFATP (SEQ ID NO:99). The underlining (amino acid residues 204-212 of mtFATP) indicates an AMP binding motif which is found in many classes of proteins; the underlining at amino acid residues 204-507 of the mtFATP sequence indicates the FATP 360 amino acid signature sequence.

-7-

Figures 2A-2D show results of LCFA uptake assays. Figures 2A-2D: COS cells were cotransfected using the DEAE-dextran method with the mammalian expression vectors pCDNA-CD2 either alone (control; Figure 2A) or in combination with one of the FATP-containing expression vectors (pCDNA-mmFATP1, Figure 2B; pCDNA-mmFATP2, Figure 2C; or pCMV-SPORT2-mmFATP5, Figure 2D) as described in Materials and Methods for Example 2. COS cells were gated on forward scatter (FSC) and side scatter (SS), and the results shown represent >10,000 cells. Cells exhibiting >300 CD2 fluorescence units (vertical line) representing 15% of all cells were deemed CD2 positive.

Figure 3 is a graph of fluorescence of cells expressing a FATP gene. As in Figures 2A-2D, COS cells were cotransfected with pCDNA-CD2 either alone (control) or in combination with one of the FATP-containing expression vectors (pCDNA-mmFATP1, pCDNA-mmFATP2, pCMV-SPORT2-mmFATP5, or pCDNA-ceFATPb). The mean BODIPY-FA fluorescence of the CD2-positive cells is plotted; results shown represent the average of three experiments, each consisting of greater than 50,000 COS cells. Note that a logarithmic scale is used on the ordinate.

Figure 4 is a graph of the uptake of palmitate with time. The full-length coding region of mtFATP (squares) or a control protein (TFE3; circles) was subcloned into the inducible, prokaryotic expression vector pET (Novagen, Madison, WI). Expression from the resulting plasmid was induced (solid symbols) in transformed *E. coli* cells with 1 mM isopropyl- $\beta$ -D-thiogalactoside (IPTG) for 1 hour, or cells were left uninduced (open symbols). Data points were done in triplicate and counts were normalized to the number of bacteria as determined by OD<sub>600</sub>.

Figure 5 is a phylogenetic tree produced by aligning complete and partial sequences for *FATP* genes from human, rat, mouse, puffer fish, *D. melanogaster*, *C. elegans*, *S. cerevisiae*, and *M. tuberculosis* using ClustalX and using these data to produce a phylogenetic tree using TreeViewPPC. The bar indicates the number of substitutions per residue, i.e., 0.1 corresponds to a distance of 10 substitutions per 100 residues.

Figure 6 shows a comparison of the FATP signature sequences of mmFATP1 (SEQ ID NO:1), mmFATP5, (SEQ ID NO:2), ceFATPa (SEQ ID NO:3), scFATP (SEQ ID NO:4) and mtFATP (SEQ ID NO:5).

Figure 7 shows the sequence identity among the FATP family members and  
5 VLACs, based on the 360 amino acid signature sequence of FATP from Figure 1.

Figures 8A and 8B are the mmFATP3 DNA sequence (SEQ ID NO:6).

Figure 9 is the mmFATP3 protein sequence (SEQ ID NO:7).

Figures 10A and 10B are the mmFATP4 DNA sequence (SEQ ID NO:8).

Figure 11 is the mmFATP4 protein sequence (SEQ ID NO:9).

10 Figures 12A and 12B are the mmFATP5 DNA sequence (SEQ ID NO:10).

Figure 13 is the mmFATP5 protein sequence (SEQ ID NO:11).

Figures 14A and 14B are the hsFATP2 DNA sequence (SEQ ID NO:12).

Figure 15 is the hsFATP2 protein sequence (SEQ ID NO:13).

Figures 16A and 16B are the hsFATP3 DNA sequence (SEQ ID NO:14).

15 Figure 17 is the hsFATP3 protein sequence (SEQ ID NO:15).

Figures 18A and 18B are the hsFATP4 DNA sequence (SEQ ID NO:16).

Figure 19 is the hsFATP4 protein sequence (SEQ ID NO:17).

Figures 20A and 20B are the hsFATP5 DNA sequence (SEQ ID NO:18).

Figure 21 is the hsFATP5 protein sequence (SEQ ID NO:19).

20 Figures 22A and 22B are the hsFATP6 DNA sequence (SEQ ID NO:20).

Figure 23 is the hsFATP6 protein sequence (SEQ ID NO:21).

Figures 24A and 24B are the mtFATP DNA sequence (SEQ ID NO:22).

Figure 25 is the mtFATP protein sequence (SEQ ID NO:23).

Figure 26 shows the DNA sequence (SEQ ID NO:24) and predicted amino  
25 acid sequence (SEQ ID NO:25) of human FATP1.

Figure 27 shows the DNA sequence (SEQ ID NO:26) and predicted amino acid sequence (SEQ ID NO:27) of human FATP4.

Figure 28A is a hydrophobicity plot for hsFATP1, showing that it has multiple membrane-spanning domains.

30 Figure 28B is the amino acid composition of hsFATP1.

Figure 28C is a hydrophilicity plot for hsFATP1, made using the Kyte-Doolittle method, averaging hydrophilicity values for 18 amino acid residues at a time.

Figure 29A is a hydrophobicity plot for hsFATP4, showing that it has  
5 multiple membrane-spanning domains.

Figure 29B is a listing of the amino acid composition of hsFATP4.

Figure 29C is a hydrophilicity plot for hsFATP4, made using the Kyte-Doolittle method, averaging hydrophilicity values for 18 amino acid residues at a time.

10 Figures 30A and 30B show a comparison of the nucleotide sequence of human FATP1 (SEQ ID NO:28) and the nucleotide sequence of mouse FATP1 (SEQ ID NO:29).

Figures 31A and 31B show a comparison of the nucleotide sequence of human FATP4 (SEQ ID NO:30) and the nucleotide sequence of mouse FATP4  
15 (SEQ ID NO:31).

Figure 32 shows a comparison of the amino acid sequence of human FATP1 (SEQ ID NO:32) and the amino acid sequence of mouse FATP1 (SEQ ID NO:33). Shaded amino acid residues match the consensus sequence exactly.

Figure 33 shows a comparison at the amino acid level of human FATP4  
20 (SEQ ID NO:34) and mouse FATP4 (SEQ ID NO:35). Shaded amino acid residues match the consensus sequence exactly.

Figure 34 shows the nucleotide sequence (SEQ ID NO:36) and predicted amino acid sequence (SEQ ID NO:37) of hsFATP6.

Figure 35A is a hydrophobicity plot for hsFATP6, showing that it has  
25 multiple membrane-spanning domains.

Figure 35B is a listing of the amino acid composition of hsFATP6.

Figure 35C is a hydrophilicity plot for hsFATP6, made using the Kyte-Doolittle method, averaging hydrophilicity values for 18 amino acid residues at a time.



Figure 36 shows an alignment of the amino acid sequences of hsFATP1 (SEQ ID NO:38), hsFATP4 (SEQ ID NO:39) and hsFATP6 (SEQ ID NO:40). Shaded amino acid residues match the consensus sequence exactly.

Figure 37 shows results of assessment of fatty acid uptake by human FATP1  
5 and human FATP4. The percent of CD2-positive cells exhibiting a BODIPY-fluorescence of more than 300 arbitrary units is plotted for the three different conditions tested.

Figure 38 is a graph showing uptake of tritiated oleate, with time, by 293 cells transfected with either (diamonds) a plasmid for expression of human FATP4  
10 or (squares) a control plasmid.

Figure 39 is an illustration of the amino acid sequences of human FATP4 (SEQ ID NO:41) and mouse FATP4 (SEQ ID NO:42) compared to human FATP1 (SEQ ID NO:43). Shown by underlining are the FATP consensus sequence (236-556 of hsFATP1) and the AMP-binding motif (246-254 of hsFATP1). The human  
15 FATPs were cloned by screening libraries with sequences from ESTs (expressed sequence tags). Mouse FATP4 was cloned by PCR using degenerate primers.

Figure 40 is a graph showing the uptake, with time, of tritiated oleate by mouse enterocytes in the presence of no oligonucleotide (squares), sense oligonucleotide (circles) or antisense oligonucleotide (diamonds).

Figure 41 is a bar graph showing uptake of tritiated oleate, by mouse  
20 enterocytes in the presence of various concentrations of antisense (solid bars), mismatch (stippled bars) or sense (lined bars) oligonucleotides.

Figure 42 is a bar graph showing uptake of tritiated oleate and uptake of <sup>35</sup>S-labeled methionine by mouse enterocytes to which were added no oligonucleotide,  
25 the antisense oligonucleotide, or the mismatch oligonucleotide.

Figure 43A is the nucleotide sequence of the gene encoding mouse FATP4 (SEQ ID NO:44).

Figure 43B is the amino acid sequence of mouse FATP4 protein (SEQ ID NO:45).

Figures 44A, 44B, and 44C are the hsFATP1 DNA sequence (SEQ ID NO:46). Coding region: 175-2115 (1941 nt).  
30

Figure 45 is the hsFATP1 protein sequence (SEQ ID NO:47).

Figures 46A and 46B are the hsFATP2 DNA sequence (SEQ ID NO:48).

Coding region: 223-2085 (1863 nt).

Figure 47 is the hsFATP2 protein sequence (SEQ ID NO:49).

5 Figure 48 is the partial DNA sequence of hsFATP3 (SEQ ID NO:50).

Coding region: 1-993.

Figure 49 is the partial protein sequence of hsFATP3 (SEQ ID NO:51).

Figures 50A, 50B, and 50C are the hsFATP4 DNA sequence (SEQ ID NO:52). Coding region: 208-2139 (1932 nt).

10 Figure 51 is the hsFATP4 protein sequence (SEQ ID NO:53).

Figure 52 is the hsFATP5 partial DNA sequence (SEQ ID NO:54). Coding region: 1-1062.

Figure 53 is the hsFATP5 partial protein sequence (SEQ ID NO:55).

Figures 54A, 54B, and 54C are the hsFATP6 DNA sequence (SEQ ID NO:56). Coding region: 643-2502 (1860 nt).

Figure 55 is the hsFATP6 protein sequence (SEQ ID NO:57).

Figures 56A, 56B, and 56C are the mFATP1 DNA sequence (m=*Rattus norvegicus*; (SEQ ID NO:58). Coding region: 75-2015 (1941 nt).

Figure 57 is the mFATP1 protein sequence (SEQ ID NO:59).

20 Figures 58A, 58B, and 58C are the mFATP2 DNA sequence (SEQ ID NO:60). Coding region: 795-2657 (1863 nt).

Figure 59 is the mFATP2 protein sequence (SEQ ID NO:61).

Figures 60A and 60B are the mFATP4 partial DNA sequence (SEQ ID NO:62). Coding region: 1-1218.

25 Figure 61 is the mFATP4 partial DNA sequence (SEQ ID NO:63).

Figures 62A, 62B, and 62C are the mmFATP1 DNA sequence (SEQ ID NO:64). Coding region: 1-1944.

Figure 63 is the mmFATP1 protein sequence (SEQ ID NO:65).

Figures 64A and 64B are the mmFATP2 DNA sequence (SEQ ID NO:66).

30 Coding region: 121-1992 (1872 nt).

Figure 65 is the mmFATP2 protein sequence (SEQ ID NO:67).

Figures 66A and 66B are the mmFATP3 partial DNA sequence (SEQ ID NO:68). Coding region: 1-1830.

Figure 67 is the mmFATP3 partial protein sequence (SEQ ID NO:69).

Figures 68A, 68B, and 68C are the mmFATP4 DNA sequence (SEQ ID NO:70). Coding region: 1-1932.

Figures 69 is the mmFATP4 protein sequence (SEQ ID NO:71).

Figures 70A and 70B are the mmFATP5 DNA sequence (SEQ ID NO:72). Coding region: 60-2129.

Figure 71 is the mmFATP5 protein sequence (SEQ ID NO:73).

Figures 72A and 72B are the dmFATP partial DNA sequence (dm=*Drosophila melanogaster*; SEQ ID NO:74). Coding region: 1-1773.

Figure 73 is the dmFATP partial protein sequence (SEQ ID NO:75).

Figure 74 is the drFATP partial DNA sequence (dr=*Danio rerio*, zebrafish; SEQ ID NO:76) Coding region: 1-173.

Figure 75 is the drFATP partial protein sequence (SEQ ID NO:77).

Figure 76A and 76B are the ceFATPa DNA sequence (SEQ ID NO:78). Coding region: 1-1953.

Figure 77 is the ceFATPa protein sequence (SEQ ID NO:79).

Figures 78A and 78B are the ceFATPb DNA sequence (SEQ ID NO:80).

Coding region: 1-1968.

Figure 79 is the ceFATPb protein sequence (SEQ ID NO:81).

Figures 80A and 80B are the chFATP DNA sequence (SEQ ID NO:82; ch=*Cochliobolus heterostrophus*). Coding region: 1-1932.

Figure 81 is the chFATP protein sequence (SEQ ID NO:83).

Figure 82 is the anFATP partial protein sequence (an=*Aspergillus nidulans*; SEQ ID NO:84). Coding region: 1-597.

Figure 83 is the anFATP partial protein sequence (SEQ ID NO:85).

Figure 84 is the mgFATP partial DNA sequence (mg=*Magnaporthe grisea*, rice blast; SEQ ID NO:86). Coding region: 1-522.

Figure 85 is the mgFATP partial protein sequence (SEQ ID NO:87).

Figures 86A and 86B are the scFATP DNA sequence (SEQ ID NO:88).

Coding region: 1-1872.

Figure 87 is the scFATP protein sequence (SEQ ID NO:89).

Figures 88A and 88B are the mtFATP DNA sequence (SEQ ID NO:90).

5 Figure 89 is the mtFATP protein sequence (SEQ ID NO:91). Coding region:  
1-1794.

Figure 90 is a consensus sequence of the FATP signature sequence (SEQ ID NO:100), based on 23 independent sequences aligned in ClustalX. The height of the bar at each amino acid residue position indicates the degree of conservation at that  
10 position. Gaps have been inserted to maintain the strength of the alignment.

Figure 91 is a hydrophilicity plot for hsFATP2, made using the Kyte-Doolittle method, averaging hydrophilicity values for 18 amino acid residues at a time.

Figure 92 is a hydrophilicity plot for the hsFATP3 partial protein, made using the Kyte-Doolittle method, averaging hydrophilicity values for 18 amino acid  
15 residues at a time.

Figure 93 is a hydrophilicity plot for the hsFATP5 partial protein, made using the Kyte-Doolittle method, averaging hydrophilicity values for 18 amino acid residues at a time.

Figures 94A and 94B are a representation of the DNA sequence (SEQ ID NO:101) of the hsFATP3 gene, and the amino acid sequence (SEQ ID NO:102) of the  
20 hsFATP3 protein.

Figure 95 shows that mammalian expression constructs containing either hsFATP4 (squares and triangles) or empty control vector (circles) were stably transfected into 293 cells. Short-term uptake of Bodipy-FA in the presence of BSA  
25 was determined by FACS. The mean fluorescence of the viable cell population is expressed in arbitrary fluorescence units. FATP4 protein expression was determined by densitometry of anti-FATP4 Western blots, and is expressed in arbitrary units.

Figure 96 is a bar graph illustrating short-term uptake of Bodipy-palmitate (1  $\mu$ M), either by control cells (black bars) or FATP4-expressing cells (hatched bars),  
30 was measured in the presence of 0, 10, 100  $\mu$ M unlabeled palmitate. FA uptake was quantified by FACS and expressed in arbitrary fluorescence units.

Figure 97 shows the rate of [ $^2\text{H}$ ]palmitate uptake by 293 cells, which were stably transfected with a construct for either human FATP4 (diamonds) or an empty vector (circles), compared to that of isolated enterocytes (squares).

Figure 98 is a bar graph illustrating the results when isolated enterocytes were incubated for 48h with increasing concentrations of the FATP4 antisense oligonucleotide or with 100  $\mu\text{M}$  of a randomized control oligonucleotide with identical nucleotide composition to the FATP4 antisense oligonucleotide. The uptake of oleate by the enterocytes was then measured over a 5 min time interval (solid bars). In parallel, the levels of FATP4 protein and, as a loading control,  $\beta$ -catenin, were determined by Western blotting and quantitated using densitometry (hatched bars). FA uptake and FATP4 protein levels were normalized to that of untreated cells. The averages and standard deviations of 4 independent experiments are shown.

Figure 99 is a bar graph illustrating the uptake rates of [ $^3\text{H}$ ]oleate, [ $^3\text{H}$ ]palmitate and [ $^{35}\text{S}$ ]methionine by primary enterocytes were measured after 48h incubation with either 100  $\mu\text{M}$  FATP4 antisense (solid bars) or 100  $\mu\text{M}$  randomized control oligonucleotide (hatched bars) and expressed as % of untreated cells.

Figure 100 is a bar graph illustrating that 8 kb of FATP5 genomic sequence SEQ ID NO.: 106 is sufficient for liver specific transcription *in vitro*. A luciferase reporter construct containing 8 kb upstream of the FATP5 initiator methionine was transfected into various cell lines using calcium phosphate as described in Example 17. Forty-eight hours after transfection, luciferase activity was measured and normalized to  $\beta$ -galactosidase activity. For each cell line, fold induction was determined by dividing the relative luciferase activity of the 8 kb construct by that of the promoter-less luciferase reporter vector. The data shown represent the mean of three experiments done in triplicate. Error bars indicate the SEM.

Figure 101 is a bar graph illustrating deletion analysis of the FATP5 promoter. Constructs containing deletions of the FATP5 promoter were transfected into HepG2 cells, assayed for luciferase activity, and normalized to  $\beta$ -galactosidase (RLU). The labels on the vertical axis correspond to the length of the promoter segment as measured from the initiator methionine. The data shown represents the mean of three experiments done in triplicate. Error bars indicate the SEM.

Figure 102 is a bar graph illustrating that 271 base pairs upstream of the FATP5 initiator methionine are sufficient for liver specific luciferase activity. A luciferase reporter construct containing 271 base pairs upstream of the FATP5 initiator methionine was transfected into various cell lines using calcium phosphate as described in Methods Example 17. Forty eight hours after transfection, luciferase activity was measured and normalized to  $\beta$ -galactosidase activity. For each cell line, fold induction was determined by dividing the relative luciferase activity of the -271 base pair construct by that of the promoter-less luciferase reporter vector. The data shown represent the mean of three experiments done in triplicate. Error bars indicate the SEM.

Figures 103A and 103B illustrate mutations of the GC box which abolish transcriptional activity. A: Schematic of mutations in the GC box aligned with the normal sequence (SEQ ID NO.: 106, SEQ ID NO.: 107, SEQ ID NO.: 108). The GC box consensus sequence is underlined. B: Constructs containing 271 base pairs upstream of the FATP5 initiator methionine with the mutations in the GC box depicted in part A were transfected into HepG2 cells, assayed for luciferase activity, and normalized to  $\beta$ -galactosidase (RLU). The data shown represent the mean of three experiments done in triplicate. Error bars indicate the SEM.

Figure 104 shows a gel shift analysis of the GC box with HepG2 nuclear extracts. Schematic showing the sequence of the oligonucleotides used in gel shift studies. The numbering reflects the distance from the initiator methionine. The two pairs of oligonucleotides are indicated by the lines and labeled AF-1 (SEQ ID NO.: 111, SEQ ID NO.: 112) and AF-2 (SEQ ID NO.: 109, SEQ ID NO.: 110).

Figure 105 is a bar graph illustrating that 30bp internal deletions of the FATP5 promoter identify another region required for luciferase activity in HepG2 cells. Reporter constructs were transfected into HepG2 cells. Luciferase activity was measured and normalized to  $\beta$ -galactosidase activity (RLU). The labels on the horizontal axis correspond to the nucleotides that were deleted and the numbering on the vertical axis represents the distance from the initiator methionine. The data shown represent the mean of three experiments done in triplicate. Error bars indicate

the SEM. Note that the five fold higher RLU activity in this figure relative to Figures 101 and 103 is the result of a manufacturer change in the  $\beta$ -galactosidase reagent.

Figure 106 is a bar graph illustrating that a linker scan of the FATP5 promoter identifies two additional elements required for transcription in HepG2 cells. Reporter  
5 constructs were transfected into HepG2 cells. Luciferase activity was measured and normalized to  $\beta$ -galactosidase activity (RLU). The labels on the horizontal axis correspond to the constructs in part A. The data shown represent the mean of three experiments done in triplicate. Error bars indicate the SEM. Please note that the lower RLU activity in this figure relative to Figures 101 and 103 is also the result of a  
10 manufacturer change in the  $\beta$ -galactosidase reagent.

Figure 107 is a schematic of the FATP5 promoter (SEQ ID NO.: 113). The GC box and two motifs identified in the linker scan are boxed and labeled. An arrow indicates the translational initiator of the FATP5 protein. The two halves of the palindrome contained in the novel motifs and referred to in the discussion are  
15 underlined.

Figure 108 is a photograph showing FATP2 expression in the mouse gall bladder epithelium.

Figure 109 is a photograph showing FATP2 expression in chimpanzee liver.

Figure 110 is a photograph showing FATP5 expression in chimpanzee liver.

20 Figures 111A and 111B represent the DNA sequence (SEQ ID NO:116) of human FATP3.

Figure 112 represents the amino acid sequence (SEQ ID NO:117) of human FATP3.

Figure 113 is a bar graph showing the results of an experiment comparing  
25 fatty acid transport between cells transfected with SEQ ID NO: 116 and untransfected cells.

Figures 114A, 114B, 114C and 114D represent portions of the amino acid sequence of mmFATP4 which were produced as fusion polypeptides in *E. coli* cells.

Figure 115 is a schematic illustrating certain components of the fusion  
30 polypeptides depicted in Figures 114A-D. The schematic shows the lipocalin domain

as well as other identified motifs and notes the relative location of each in the mmFATP4 fusion polypeptide.

Figure 116 is a bar graph illustrating the results of an experiment comparing the binding capabilities of the fusion polypeptides shown in Figures 114A-D for an  
5 oleate fatty acid.

Figure 117 is a bar graph showing the results of an experiment comparing binding of various fatty acids between two of the fusion polypeptides depicted in Figure 114A-D.

Figure 118A-G illustrates the consensus sequence of hsFATP1, hsFATP2,  
10 hsFATP3, hsFATP4, hsFATP5 and hsFATP6 with the lipocalin domain and AMP-binding domain of each noted.

The foregoing and other objects, features and advantages of the invention will be apparent from the following more particular description of preferred embodiments of the invention, as illustrated in the accompanying drawings in which like reference  
15 characters refer to the same parts throughout the different views. The drawings are not necessarily to scale, emphasis instead being placed upon illustrating the principles of the invention.

#### DETAILED DESCRIPTION OF THE INVENTION

As described herein, FATPs are a large evolutionarily conserved family of  
20 proteins that mediate the transport of LCFAs into cells. The family includes proteins which are conserved from mycobacteria to humans and exhibit very different expression patterns in tissues. Specific embodiments described include FATPs from mice, humans, nematodes, fungi and mycobacteria which have been shown to be functional LCFA transporters. The term "fatty acid transport proteins" ("FATPs") as  
25 used herein, refers to the proteins described herein as FATP1, FATP2, FATP3, FATP4, FATP5 and FATP6, which have been described in one or more species of mammals, as well as mtFATP, ceFATP, scFATP, anFATP, mgFATP, and chFATP, and other proteins sharing at least about 50% amino acid sequence similarity,



preferably at least about 60% sequence similarity, more preferably at least about 70% sequence similarity, and still more preferably, at least about 80% sequence similarity, and most preferably, at least about 90% sequence similarity in the approximately 360 amino acid signature sequence. The approximately 360 amino acid FATP signature sequence is shown in Figure 1. The consensus sequence of the signature sequence is shown in Figure 90. The nomenclature used herein to refer to FATPs includes a species-specific prefix (e.g., mm, *Mus musculus*; hs or h, *Homo sapiens* or human; mt *M. tuberculosis*; dm, *D. melanogaster*; ce, *C. elegans*; sc, *Saccharomyces cerevisiae*) and a number such that mammalian homologues in different species share the same number. For example, six human and five mouse *FATP* genes which are expressed in a variety of tissues are described herein and are referred to, respectively, as hsFATP1-hsFATP6 and mmFATP1-mmFATP5; for example, hsFATP4 and mmFATP4 are the human and mouse orthologs.

Expression patterns of human and mouse FATPs have been assessed and are described below. Briefly, results of these assessments show that FATP5 is a liver-specific gene. FATP2 is highly expressed in liver, kidney and gall bladder epithelium. Both of these proteins, as well as FATP4 and FATPs from nematodes and mycobacteria, have been shown to be functional LCFA transporters. Results have also shown that FATP4 mRNA is present at high levels in epithelial cells of two regions of the small intestine (the jejunum and ileum) and at lower, but significant, levels in a third region (the duodenum). They further showed that FATP2 mRNA is present in epithelial cells of the duodenum at a level similar to that of FATP4 mRNA levels, but is present at lower levels in the jejunum and ileum. FATP4 mRNA was absent from other cell types of the small intestine and no FATP4 mRNA could be detected in any cells of the colon. No signals above background could be detected for FATP1, FATP3 and FATP5 in any of the intestinal tissues. Thus, FATP4 is the major FATP in the mouse small intestine, which supports a major role for FATP4 (along with FATP2 to a lesser extent) in absorption of free fatty acids. hsFATP4 was clearly expressed in the jejunum and ileum; expression was absent in the stomach. This, too, is consistent with a major role for FATP4 in absorption of fatty acids in the human gut. Analysis of FATP expression in human tissues, also described in detail

below, showed that hsFATP6, which has no mouse ortholog as yet, is expressed at high levels in the heart and at low levels in the placenta, but is undetectable in the other tissues assessed (Example 9). This is consistent with a major role for FATP6 in absorption of fatty acids in the heart.

5           Analysis of FATP3 expression in murine tissues, also described in detail below, showed that expression occurs at detectable levels in liver, spleen, heart, kidney, testis, white adipose tissue, exocrine and endocrine pancreatic cells, and also in lung tissues. FATP3 is expressed at high levels in type-II pneumocytes, a cell type noted for secretion a surfactant, a phospholipid-rich film critical for lung function  
10 (Example 19).

Long chain fatty acids (LCFAs) are an important energy source for pro- and eukaryotes and are involved in diverse cellular processes, such as membrane synthesis, intracellular signaling, protein modification, and transcriptional regulation. In developed Western countries, human dietary lipids are mainly di- and triglycerides  
15 and account for approximately 40% of caloric intake (Weisburger, J. H. (1997) *J. Am. Diet. Assoc.* 97:S16-S23). These lipids are broken down into fatty acids and glycerol by pancreatic lipases in the small intestine (Chapus, C., Rovey, M., Sarda, L & Verger, R. (1988) *Biochimie* 70:1223-34); LCFAs are then transported into brush border cells, where the majority is re-esterified and secreted into the lymphatic system  
20 as chylomicrons (Green, P.H. & Riley, J.W. (1981) *Aust. N.Z.J. Med.* 11:84-90). Fatty acids are liberated from lipoproteins by the enzyme lipoprotein lipase, which is bound to the luminal side of endothelial cells (Scow, R.O. & Blachette-Mackie, E.J. (1992) *Mol. Cell. Biochem* 116:181-191). "Free" fatty acids in the circulation are bound to serum albumin (Spector, A.A. (1984) *Clin. Physiol. Biochem* 2:123-134)  
25 and are rapidly incorporated by adipocytes, hepatocytes, and cardiac muscle cells. The latter derive 60-90% of their energy through the oxidation of LCFAs (Neely, J.F. Rovetto, M.J. & Oram, J.F. (1972) *Prog. Cardiovasc. Dis.* 15:289-329). Although saturable and specific uptake of LCFAs has been demonstrated for intestinal cells, hepatocytes, cardiac myocytes, and adipocytes, the molecular mechanisms of LCFA  
30 transport across the plasma membrane have remained controversial (Hui, T.Y. & Bernlohr, D.A. (1997) *Front. Biosci.* 15:d222-31-d231; Schaffer, J.E. & Lodish, H.F,

(1995) *Trends Cardiovasc. Med.* 5:218-224). Described herein is a large family of highly homologous mammalian LCFA transporters which show wide expression, including in all tissues relevant to fatty acid metabolism. Further described are novel members of this family in other species, including mycobacterial and nematode  
5 FATPs which, like their mammalian counterparts, are functional fatty acid transporters.

The discovery of a diverse but highly homologous family of FATPs is reminiscent of the glucose transporter family. In a manner similar to the FATPs, the glucose transporters have very divergent patterns of tissue expression (McGowan,  
10 K.M., Long, S.D. & Pekala, P.H. (1995) *Pharmacol. Ther.* 66:465-505). The FATPs, like glucose transporters, may also differ in their substrate specificities, uptake kinetics, and hormonal regulation (Thorens, B. (1996) *Am. J. Physiol.* 270:G541-G553). Indeed, the levels of fatty acids in the blood, like those of glucose, can be regulated by insulin and are dysregulated in diseases such as noninsulin-dependent  
15 diabetes and obesity (Boden, G. (1997) *Diabetes* 46:3-10). The underlying mechanisms for the regulation of free fatty acid concentrations in the blood are not understood, but could be explained by hormonal modulation of FATPs.

Insulin-resistance is thought to be the major defect in non insulin-dependent diabetes mellitus (NIDDM) and is one of the earliest manifestations of NIDDM  
20 (McGarry (1992) *Science* 258:766-770). Free fatty acids (FFAs) may provide an explanation for why obesity is a risk factor for NIDDM. Plasma levels of FFAs are elevated in diabetic patients (Reaven *et al.* (1988) *Diabetes* 37:1020). Elevated plasma free fatty acids (FFAs) have been demonstrated to induce insulin-resistance in whole animals and humans (Boden (1998) *Front. Biosci.* 3:D169-D175). This  
25 insulin-resistance is likely mediated by effects of FFAs on a variety of issues. FFAs added to adipocytes *in vitro* induce insulin resistance in this cell type as evidenced by inhibition of insulin-induced glucose transport (Van Epps-Fung *et al.* (1997) *Endocrinology* 138:4338-4345). Rats fed a high fat diet developed skeletal muscle insulin resistance as evidenced by a decrease in insulin-induced glucose uptake by  
30 skeletal muscle (Han *et al.*, (1997) *Diabetes* 46:1761-1767). In addition, elevated plasma FFAs increase insulin-suppressed endogenous glucose production in the liver

(Boden (1998) *Front. Biosci.* 3:D169-D175), thus increasing hepatic glucose output. It has been postulated that the adverse effects of plasma free fatty acids are due to the FFAs being taken up into the cell, leading to an increase in intracellular long chain fatty acyl CoA; intracellular long chain acyl CoAs are thought to mediate the effects of FFAs inside the cell. Thus, fatty acid induced insulin-resistance may be prevented by blocking uptake of FFAs into select tissues, in particular liver (by blocking FATP2 and/or FATP5), adipocyte (by blocking FATP1), and skeletal muscle (by blocking FATP1). Blocking intestinal fat absorption (by blocking FATP4) is also expected to reduce plasma FFA levels and thus improve insulin resistance.

During the pathogenesis of NIDDM insulin-resistance can initially be counteracted by increasing insulin output by the pancreatic beta cell. Ultimately, this compensation fails, beta cell function decreases and overt diabetes results (McGarry (1992) *Science* 258: 766-770). Manipulating beta cell function is a second point where fatty acid transporter blockers may be beneficial for diabetes. While no FATP homolog has been identified so far that is expressed in the beta cell of the pancreas, the data described below suggest the existence of such a transporter and the sequence information included herein provides the means to identify such a transporter by degenerate PCR, using primers to regions conserved in all FATP family members or by low stringency hybridization. It has been demonstrated that exposure of pancreatic beta-cells to FFAs increases the basal rate of insulin secretion; this in turn leads to a decrease in the intracellular stores of insulin, resulting in decreased capacity for insulin secretion after chronic exposure (Bollheimer *et al.*, (1998) *J. Clin. Invest.* 101:1094-1101). The effects of FFAs are again likely to be mediated by intracellular long chain fatty acyl CoA molecules (Liu *et al.*, (1998) *J. Clin. Invest.* 101:1870-1875). FFAs have also been demonstrated to increase beta cell apoptosis (Shimabukuro *et al.*, (1998) *Proc. Nat. Acad. Sci. USA* 95:2498-2502), possibly contributing to the decrease in beta cell numbers in late stage NIDDM.

Another finding with potentially broad implications is the identification of a FATP homologue in *M. tuberculosis*. Tuberculosis causes more deaths worldwide than any other infectious agent and drug-resistant tuberculosis is re-emerging as a problem in industrialized nations (Bloom, B.R. & Small, P.M. (1998) *N. Engl. J.*

*Med.* 338:677-678). *Mycobacterium tuberculosis* has about 250 enzymes involved in fatty acid metabolism, compared with only about 50 in *E. coli*. It has been suggested that, living as a pathogen, the mycobacteria are largely lipolytic, rather than lipogenic, relying on the lipids within mammalian cells and the tubercle (Cole, S.T. *et al.*,  
5 *Nature* 393:537-544 (1998)). The *de novo* synthesis of fatty acids in *Mycobacterium leprae* is insufficient to maintain growth (Wheeler, P.R., Bulmer, K & Ratledge, C. (1990) *J. Gene. Microbiol.* 136:211-217). Thus, it is reasonable to expect that inhibitors of mtFATP will serve as therapeutics for tuberculosis. FATPs expressed in mycobacteria can be targeted to reduce or prevent replication of mycobacteria (e.g., to  
10 reduce or prevent replication of *M. tuberculosis*) and, thus, reduce or prevent their adverse effects. For example, a FATP or FATPs expressed by *M. tuberculosis* can be targeted and inhibited, thus reducing or preventing growth of this pathogen (and tuberculosis in humans and other mammals). An inhibitor of an *M. tuberculosis* FATP can be identified, using methods described herein (e.g., expressing the FATP  
15 in an appropriate host cell, such as *E. coli* or COS cells; contacting the cells with an agent or drug to be assessed for its ability to inhibit the FATP and, as a result, mycobacterial growth, and assessing its effects on growth). A drug or agent identified in this manner can be further tested for its ability to inhibit a *M. tuberculosis* FATP and *M. tuberculosis* infection in an appropriate animal model or  
20 in humans. A method of inhibiting mycobacterial growth, particularly growth of *M. tuberculosis*, and compounds useful as drugs for doing so are also the subject of this invention.

An isolated polynucleotide encoding mtFATP, like other polynucleotides encoding FATPs of the FATP family, can be incorporated into vectors, nucleic acids  
25 of viruses, and other nucleic acid constructs that can be used in various types of host cells to produce mtFATP. This mtFATP can be used, as it appears on the surface of cells, or in various artificial membrane systems, to assess fatty acid transport function, to identify ligands and molecules that are modulators of fatty acid transport activity. Molecules found to be inhibitors of mtFATP function can be incorporated  
30 into pharmaceutical compositions to administer to a human for the treatment of tuberculosis.

Particular embodiments of the invention are polynucleotides encoding a FATP of *Cochliobolus (Helminthosporium) heterostrophus* or portions or variants thereof, the isolated or recombinantly produced FATP, methods for assessing whether an agent binds to the chFATP, and further methods for assessing the effect of an agent being tested for its ability to modulate fatty acid transport activity.

*Cochliobolus heterostrophus* is an ascomycete that is the cause of southern corn leaf blight, an economically important threat to the corn crop in the United States. The related species *C. sativus* causes crown rot and common root rot in wheat and barley. One or more FATPs of *C. heterostrophus* can be targeted for the identification of an inhibitor of chFATP function, which can be then be used as an agent effective against infection of plants by *C. heterostrophus* and related organisms. Methods described herein that were applied in studying the expression of a FATP gene and the function of the FATP in its natural site of expression or in a host cell, can be used in the study of the chFATP gene and protein.

*Magnaporthe grisea* (rice blast) is an economically important fungal pathogen of rice. Further embodiments of the invention are nucleic acid molecules encoding a FATP of *Magnaporthe grisea*, portions thereof, or variants thereof, isolated mgFATP, nucleic acid constructs, and engineered cells expressing mgFATP. Other aspects of the invention are assays to identify an agent which binds to mgFATP and assays to identify an agent which modulates the function of mgFATP in cells in which mgFATP is expressed or in artificial membrane systems. Agents identified as inhibiting mgFATP activity can be developed into anti-fungal agents to be used to treat rice infected with rice blast.

*Caenorhabditis elegans* is a nematode related to plant pathogens and human parasites. An isolated polynucleotide which encodes ceFATP, like other polynucleotides encoding FATPs of the FATP family described herein, can be incorporated into nucleic acid vectors and other constructs that can be used in various types of cells to produce ceFATP. ceFATP as it occurs in cells or as it can be isolated or incorporated into various artificial or reconstructed membrane systems, can be used to assess fatty acid transport, and to identify ligands and agents that modulate fatty acid transport activity. Agents found by such assays to be inhibitors of ceFATP

activity can be incorporated into compositions for the treatment of diseases caused by genetically related organisms with a FATP of similar sensitivity to the agents.

*Aspergillus nidulans* is one of a family of fungal species that can infect humans. Further embodiments of the invention of the family of polynucleotides  
5 encoding FATPs are polynucleotides encoding a FATP of *Aspergillus nidulans*, and vectors and host cells that can be constructed to comprise such polynucleotides. Further embodiments are a polypeptide encoded by such polynucleotides, portions thereof having one or more functions characteristic of a FATP, and various methods. The methods include those for identifying agents that bind to a FATP and those for  
10 assessing the effect of an agent being tested for its ability to modulate fatty acid transport activity. Those agents found to inhibit fatty acid transport function can be used in compositions as anti-fungal pharmaceuticals, or can be modified for greater effectiveness as a pharmaceutical.

One aspect of the invention relates to isolated nucleic acids that encode a  
15 FATP as described herein, such as those FATPs having an amino acid sequence in Figure 45 (SEQ ID NO:47), Figure 47 (SEQ ID NO:49), Figure 112 (SEQ ID NO:117), Figure 51 (SEQ ID NO:53), Figure 53 (SEQ ID NO:55), and Figure 55 (SEQ ID NO:57) and nucleic acids closely related thereto as described herein.

Using the information provided herein, such as a nucleic acid sequence set  
20 forth in Figures 44A-44C (SEQ ID NO:46), Figures 46A and 46B (SEQ ID NO:48), Figure 112 (SEQ ID NO:116), Figures 50A-50C (SEQ ID NO:52), Figure 52 (SEQ ID NO:54), and Figures 54A-54C (SEQ ID NO:56), a nucleic acid of the invention encoding a FATP polypeptide has been obtained using standard cloning and screening methods, such as those for cloning and sequencing cDNA library  
25 fragments, followed by obtaining a full length clone. For example, to obtain a nucleic acid of the invention, a library of clones of cDNA of human or other mammalian DNA can be probed with a labeled oligonucleotide, such as a radiolabeled oligonucleotide, preferably about 17 nucleotides or longer, derived from a partial sequence. Clones carrying DNA identical to that of the probe can then be  
30 distinguished using stringent (also, "high stringency") hybridization conditions. By sequencing the individual clones thus identified with sequencing primers designed

from the original sequence it is then possible to extend the sequence in both directions to determine the full length sequence. Suitable techniques are described, for example, in *Current Protocols in Molecular Biology* (F.M. Ausubel et al, eds), containing supplements through Supplement 42, 1998, John Wiley and Sons, Inc.,  
5 especially chapters 5, 6 and 7.

Embodiments of the invention include isolated nucleic acid molecules comprising any of the following nucleotide sequences: 1.) a nucleotide sequence which encodes a protein comprising the amino acid sequence of hsFATP1 (SEQ ID NO:47), the amino acid sequence of hsFATP2 (SEQ ID NO:49), the amino acid  
10 sequence of hsFATP3 (SEQ ID NO:117), the amino acid sequence of hsFATP4 (SEQ ID NO: 53), the amino acid sequence of hsFATP5 (SEQ ID NO:55) or the amino acid sequence of hsFATP6 (SEQ ID NO:57); 2.) nucleotide sequences of hsFATP1, hsFATP2, hsFATP3, hsFATP4, hsFATP5, or hsFATP6 (SEQ ID NO:46, 48, 116, 52, 54, or 56, respectively); 3.) a nucleotide sequence which is complementary to the  
15 nucleotide sequence of hsFATP1 (SEQ ID NO:46), hsFATP2 (SEQ ID NO:48), hsFATP3 (SEQ ID NO:116), hsFATP4 (SEQ ID NO:52), hsFATP5 (SEQ ID NO:54) or hsFATP6 (SEQ ID NO:56); 4.) a nucleotide sequence which consists of the coding region of hsFATP1 (SEQ ID NO:46), the coding region of hsFATP2 (SEQ ID NO:48), the coding region of hsFATP3 (SEQ ID NO:116), the coding region of  
20 hsFATP4 (SEQ ID NO:52), the coding region of hsFATP5 (SEQ ID NO:54), or the coding region of hsFATP6 (SEQ ID NO:56).

The invention further relates to nucleic acids (nucleic acid molecules or polynucleotides) having nucleotide sequences identical over their entire length to those shown in the figures, for instance Figures 44A-44C (SEQ ID NO:46), Figures  
25 46A and 46B (SEQ ID NO:48), Figures 111A-B (SEQ ID NO:116), Figures 50A-50C (SEQ ID NO:52), Figure 52 (SEQ ID NO:54), and Figures 54A-54C (SEQ ID NO:56). It further relates to DNA, which due to the degeneracy of the genetic code, encodes a FATP encoded by one of the FATP-encoding DNAs, whose amino acid sequence is provided herein. Also provided by the invention are nucleic acids having  
30 the coding sequences for the mature polypeptides or fragments in reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-,



or pro- or prepro- protein sequence. The nucleic acids of the invention encompass nucleic acids that include a single continuous region or discontinuous regions encoding the polypeptide, together with additional regions, that may also contain coding or non-coding sequences. The nucleic acids may also contain non-coding  
5 sequences, including, for example, but not limited to, non-coding 5' and 3' sequences, such as the transcribed, non-translated sequences, termination signals, ribosome binding sites, sequences that stabilize mRNA, introns, polyadenylation signals, and additional coding sequences which encode additional amino acids. For example, a marker sequence that facilitates purification of the fused polypeptide can be encoded.  
10 In certain embodiments of the invention, the marker sequence can be a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc., Venlo, The Netherlands) and described in Gentz *et al.*, *Proc. Natl. Acad. Sci. USA* 86: 821-824 (1989), or an HA tag (Wilson *et al.*, *Cell* 37: 767 (1984)), or a sequence encoding glutathione S-transferase of *Schistosoma japonicum* (vectors available from Pharmacia; see Smith,  
15 D.B. and Johnson K.S., *Gene* 67:31 (1988) and Kaelin, W.G. *et al.*, *Cell* 70:351 (1992)). Nucleic acids of the invention also include, but are not limited to, nucleic acids comprising a structural gene and its naturally associated sequences that control gene expression.

The invention further relates to nucleic acids (nucleic acid molecules or  
20 polynucleotides) that encode a FATP polypeptide. In a particular embodiment, a nucleic acid encodes a portion of a FATP which includes a motif or domain, for example, a lipocalin domain or an AMP-binding domain. Such a polypeptide portion can be a functional portion of a FATP protein. The term "lipocalin domain" is an art recognized term and as used herein refers to a particular domain present in FATP  
25 proteins. This domain is described as including regions of sequence homology as well as a common tertiary structure represented as an eight stranded antiparallel beta-barrel. (see Banaszak, L. *et al.*, *Advances in Protein Chemistry*, 45: 89-151). Many lipocalin domains can be identified structurally as a sequence contained within the general formula: [DENG]-X-[DENQGSTARK]-X(0,2)-[DENQARK]-[LIVFY]-  
30 {CP}-G-{C}-W-[FYWLRH-X]-[LIVMTA], *e.g.*, the lipocalin signature sequence or consensus pattern (SEQ ID NO: 125). One skilled in the art will recognize that a

-27-

lipocalin domain for a particular FATP protein can vary in sequence from this general formula. A FATP lipocalin domain can be, for example, identical to the lipocalin signature sequence or can exhibit 60, 65, 70, 75, 80, 85, 90, 95 or greater per cent sequence identity compared to the general formula provided that it retains lipocalin binding function. For example, a lipocalin domain for each of the human FATPs, hsFATP1 (SEQ ID NO: 126), hsFATP2 (SEQ ID NO: 127), hsFATP3 (SEQ ID NO: 128), hsFATP4 (SEQ ID NO: 129), hsFATP5 (SEQ ID NO: 130), and hsFATP6 (SEQ ID NO: 131) has been identified. The pattern of these lipocalin domains are highly conserved across the FATP family.

10           A nucleic acid encoding a portion of a FATP polypeptide can encode one or more domains, and also can include additional nucleotides. For example, the nucleic acid can also include nucleotide sequences that encode a portion of a FATP protein that is upstream from a lipocalin domain. As the term "upstream" or "upstream sequences" is used herein in relation to the lipocalin domain, it is intended to refer to the nucleotide sequence which encodes all or a portion of a FATP protein located between the signal peptide (when one is present) and the lipocalin domain. In the absence of a signal peptide, the term refers to the nucleotide sequence which encodes all or a portion of a FATP protein between the lipocalin domain and the amino terminus (see Figure 115).

20           The invention further relates to variants, including naturally-occurring allelic variants, of those nucleic acids described specifically herein by DNA sequence, that encode variants of such polypeptides as those having the amino acid sequences shown in Figure 45 (SEQ ID NO:47), Figure 47 (SEQ ID NO:49), Figure 112 (SEQ ID NO:117), Figure 51 (SEQ ID NO:53) Figure 53 (SEQ ID NO:55), or Figure 55 (SEQ ID NO:57). Such variants include nucleic acids encoding variants of the above-listed amino acid sequences, wherein those variants have several, such as 5 to 10, 1 to 5, or 3, 2 or 1 amino acids substituted, deleted, or added, in any combination. Variants include polynucleotides encoding polypeptides with at least 95% but less than 100% amino acid sequence identity to the polypeptides described herein by amino acid sequence. Variant polynucleotides hybridize, under low to high stringency conditions, to the alleles described herein by DNA sequence. In one

embodiment, variants have silent substitutions, additions and deletions that do not alter the properties and activities of the FATP. Allelic variants of the polynucleotides encoding hsFATP1 (Figure 45; SEQ ID NO:47), hsFATP2 (Figure 47; SEQ ID NO:49), hsFATP3 (Figure 112; SEQ ID NO:117), hsFATP4 (Figure 51; SEQ ID NO:53), hsFATP5 (Figure 53; SEQ ID NO:55) and hsFATP6 (Figure 55; SEQ ID NO:57) will be identified as mapping to chromosomal locations listed for the corresponding wild type genes in Table 2 in Example 1.

Orthologous genes are gene loci in different species that are sufficiently similar to each other in their nucleotide sequences to suggest that they originated from a common ancestral gene. Orthologous genes arise when a lineage splits into two species, rather than when a gene is duplicated within a genome. Proteins that are orthologs are encoded by genes of two different species, wherein the genes are said to be orthologous.

The invention further relates to polynucleotides encoding polypeptides which are orthologous to those polypeptides having a specific amino acid sequence described herein, such as the amino acid sequences shown in Figure 45 (SEQ ID NO:47), Figure 47 (SEQ ID NO:49), Figure 112 (SEQ ID NO:117), Figure 51 (SEQ ID NO:53), Figure 53 (SEQ ID NO:55), or Figure 55 (SEQ ID NO:57). These polynucleotides, which can be called ortholog polynucleotides, encode orthologous polypeptides that can range in amino acid sequence identity to a reference amino acid sequence described herein, from about 65% to less than 100%, but preferably 70% to 80%, more preferably 80% to 90%, and still more preferably 90% to less than 100%. Orthologous polypeptides can also be those polypeptides that range in amino acid sequence similarity to a reference amino acid sequence described herein from about 75% to 100%, within the signature sequence. The amino acid sequence similarity between the signature sequences of orthologous polypeptides is preferably 80%, more preferably 90%, and still more preferably, 95%. The ortholog polynucleotides encode polypeptides that have similar functional characteristics (e.g., fatty acid transport activity) and similar tissue distribution, as appropriate to the organism from which the ortholog polynucleotides can be isolated.

Ortholog polynucleotides can be isolated from (e.g., by cloning or nucleic acid amplification methods) a great number of species, as shown by the sample of FATPs from evolutionarily divergent species described herein (see, e.g., Figures 44A-C through Figure 89). Ortholog polynucleotides corresponding to those in Figure 45  
5 (SEQ ID NO:47), Figure 47 (SEQ ID NO:49), Figures 111A-B (SEQ ID NO:116), Figure 51 (SEQ ID NO:53), Figure 52 (SEQ ID NO:55) and Figure 55 (SEQ ID NO:57) are those which can be isolated from mammals such as rat, dog, chimpanzee, monkey, baboon, pig, rabbit and guinea pig, for example.

Further variants that are fragments of the nucleic acids of the invention may  
10 be used to synthesize full-length nucleic acids of the invention, such as by use as primers in a polymerase chain reaction. As used herein, the term primer refers to a single-stranded oligonucleotide which acts as a point of initiation of template-directed DNA synthesis under appropriate conditions (e.g., in the presence of four different nucleoside triphosphates and an agent for polymerization, such as DNA or  
15 RNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. The appropriate length of a primer depends on the intended use of the primer, but typically ranges from 15 to 30 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template, but  
20 must be sufficiently complementary to hybridize with a template. The term primer site refers to the area of the target DNA to which a primer hybridizes. The term primer pair refers to a set of primers including a 5' (upstream) primer that hybridizes with the 5' end of the DNA sequence to be amplified and a 3' (downstream) primer that hybridizes with the complement of the 3' end of the sequence to be amplified.

25 Further embodiments of the invention are nucleic acids that are at least 80% identical over their entire length to a nucleic acid described herein, for example a nucleic acid having the nucleotide sequence in Figures 44A-44C (SEQ ID NO:46), Figures 46A-46B (SEQ ID NO:48), Figures 111A-B (SEQ ID NO:116), Figures 50A-50C (SEQ ID NO:52), Figure 52 (SEQ ID NO:54), and Figures 54A-54C (SEQ ID  
30 NO:56). Additional embodiments are nucleic acids, and the complements of such nucleic acids, having at least 90% nucleotide sequence identity to the above-

described sequences, and nucleic acids having at least 95% nucleotide sequence identity. In preferred embodiments, DNA of the present invention has 97% nucleotide sequence identity, 98% nucleotide sequence identity, or at least 99% nucleotide sequence identity with the DNA whose sequences are presented herein.

5           Other embodiments of the invention are nucleic acids that are at least 80% identical in nucleotide sequence to a nucleic acid encoding a polypeptide having an amino acid sequence as set forth in Figure 45 (SEQ ID NO:47), Figure 47 (SEQ ID NO:49), Figure 112 (SEQ ID NO:117), Figure 51 (SEQ ID NO:53), Figure 53 (SEQ ID NO:55) or Figure 55 (SEQ ID NO:57), or as such amino acid sequences are set  
10   forth elsewhere herein, and nucleic acids that are complementary to such nucleic acids. Specific embodiments are nucleic acids having at least 90% nucleotide sequence identity to a nucleic acid encoding a polypeptide having an amino acid sequence as described in the list above, nucleic acids having at least 95% sequence identity, and nucleic acids having at least 97% sequence identity.

15           The terms "complementary" or "complementarity" as used herein, refer to the natural binding of polynucleotides under permissive salt and temperature conditions by base-pairing. Complementarity between two single-stranded molecules may be "partial" in which only some of the nucleic acids bind, or it may be complete when total complementarity exists between the single-stranded molecules (that is, when A-  
20   T and G-C base pairing is 100% complete). The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, which depend on binding between nucleic acid strands.

          The invention further includes nucleic acids that hybridize to the above-  
25   described nucleic acids, especially those nucleic acids that hybridize under stringent hybridization conditions. "Stringent hybridization conditions" or "high stringency conditions" generally occur within a range from about  $T_m$  minus 5°C (5° C below the strand dissociation temperature or melting temperature ( $T_m$ ) of the probe nucleic acid molecule) to about 20° C to 25° C below  $T_m$ . As will be understood by those of skill  
30   in the art, the stringency of hybridization may be altered in order to identify or detect molecules having identical or related polynucleotide sequences. An example of high

stringency hybridization follows. Hybridization solution is (6x SSC/10 mM EDTA/0.5% SDS/5x Denhardt's solution/100 µg/ml sheared and denatured salmon sperm DNA). Hybridization is at 64-65°C for 16 hours. The hybridized blot is washed two times with 2x SSC/0.5% SDS solution at room temperature for 15 minutes each, and two times with 0.2x SSC/0.5% SDS at 65°C, for one hour each. Further examples of high stringency conditions can be found on pages 2.10.1-2.10.16 (see particularly 2.10.8-11) and pages 6.3.1-6 in *Current Protocols in Molecular Biology* (Ausubel, F.M. *et al.*, eds., containing supplements up through Supplement 42, 1998). Examples of high, medium, and low stringency conditions can be found on pages 36 and 37 of WO 98/40404, which are incorporated herein by reference.

The invention further relates to nucleic acids obtainable by screening an appropriate library with a probe having a nucleotide sequence such as that set forth in Figures 44A-44C (SEQ ID NO:46), Figures 46A-46B (SEQ ID NO:48), Figure 111 (SEQ ID NO:116), Figures 50A-50C (SEQ ID NO:52), Figure 52 (SEQ ID NO:54) or Figures 54A-54C (SEQ ID NO:56), or a probe which is a sufficiently long fragment of any of the above; and isolating the nucleic acid. Such probes generally can comprise at least 15 nucleotides. Nucleic acids obtainable by such screenings may include RNAs, cDNAs and genomic DNA, for example, encoding FATPs of the FATP family described herein.

Further uses for the nucleic acid molecules of the invention, whether encoding a full-length FATP or whether comprising a contiguous portion of a nucleic acid molecule such as one given in SEQ ID NO:46, 48, 116, 52, 54, or 56, include use as markers for tissues in which the corresponding protein is preferentially expressed (to identify constitutively expressed proteins or proteins produced at a particular stage of tissue differentiation or stage of development of a disease state); as molecular weight markers on southern gels; as chromosome markers or tags (when labeled, for example with biotin, a radioactive label or a fluorescent label) to identify chromosomes or to map related gene positions; to compare with endogenous DNA sequences in a mammal to identify potential genetic disorders; as probes to hybridize and thus identify, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; as a probe to "subtract-out" known sequences in the

-32-

process of discovering other novel nucleic acid molecules; for selecting and making oligomers for attachment to a "gene chip" or other support, to be used, for example, for examination of expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or to elicit  
5 another immune response.

In certain embodiments, a contiguous portion can be about 15, 25, 30, 40, 50, 75, 100, 200, 300, 400, 500, 750, 1000, 1100, 1250, 1500 or more nucleotides in length. In a particular embodiment, the contiguous portion encompasses the signature sequence of a FATP and is about 1080 nucleotides in length.

10 Further methods to obtain nucleic acids encoding FATPs of the FATP family include PCR and variations thereof (e.g., "RACE" PCR and semi-specific PCR methods). Portions of the nucleic acids having a nucleotide sequence set forth in Figures 44A-44C (SEQ ID NO:46), Figures 46A-46B (SEQ ID NO:48), Figures 111A-B (SEQ ID NO:116), Figures 50A-50C (SEQ ID NO:52), Figure 52 (SEQ ID  
15 NO:54) or Figures 54A-54C (SEQ ID NO:56), (especially "flanking sequences" on either side of a coding region) can be used as primers in methods using the polymerase chain reaction, to produce DNA from an appropriate template nucleic acid.

Once a fragment of the FATP gene is generated by PCR, it can be sequenced,  
20 and the sequence of the product can be compared to other DNA sequences, for example, by using the BLAST Network Service at the National Center for Biotechnology Information. The boundaries of the open reading frame can then be identified using semi-specific PCR or other suitable methods such as library screening. Once the 5' initiator methionine codon and the 3' stop codon have been  
25 identified, a PCR product encoding the full-length gene can be generated using genomic DNA as a template, with primers complementary to the extreme 5' and 3' ends of the gene or to their flanking sequences. The full-length genes can then be cloned into expression vectors for the production of functional proteins.

The invention also relates to isolated proteins or polypeptides such as those  
30 encoded by nucleic acids of the present invention. Isolated proteins can be purified from a natural source or can be made recombinantly. Proteins or polypeptides

referred to herein as "isolated" are proteins or polypeptides that exist in a state different from the state in which they exist in cells in which they are normally expressed in an organism, and include proteins or polypeptides obtained by methods described herein, similar methods or other suitable methods, and also include

5 essentially pure proteins or polypeptides, proteins or polypeptides produced by chemical synthesis or by combinations of biological and chemical methods, and recombinant proteins or polypeptides which are isolated. Thus, the term "isolated" as used herein, indicates that the polypeptide in question exists in a physical milieu distinct from that in which it occurs in nature. Thus, "isolated" includes existing in

10 membrane fragments and vesicles membrane fractions, liposomes, lipid bilayers and other artificial membrane systems. An isolated FATP may be substantially isolated with respect to the complex cellular milieu in which it naturally occurs, and may even be purified essentially to homogeneity, for example as determined by PAGE or column chromatography (for example, HPLC), but may also have further cofactors or

15 molecular stabilizers, such as detergents, added to the purified protein to enhance activity. In one embodiment, proteins or polypeptides are isolated to a state at least about 75% pure; more preferably at least about 85% pure, and still more preferably at least about 95% pure, as determined by Coomassie blue staining of proteins on SDS-polyacrylamide gels. Proteins or polypeptides referred to herein as "recombinant" are

20 proteins or polypeptides produced by the expression of recombinant nucleic acids.

In a preferred embodiment, an isolated polypeptide comprising a FATP, a functional portion thereof, or a functional equivalent of the FATP, has at least one function characteristic of a FATP, for example, transport activity, binding function (e.g., a domain which binds to AMP), or antigenic function (e.g., binding of

25 antibodies that also bind to a naturally-occurring FATP, as that function is found in an antigenic determinant). Functional equivalents can have activities that are quantitatively similar to, greater than, or less than, the reference protein. These proteins include, for example, naturally occurring FATPs that can be purified from tissues in which they are produced (including polymorphic or allelic variants),

30 variants (e.g., mutants) of those proteins and/or portions thereof. Such variants include mutants differing by the addition, deletion or substitution of one or more



amino acid residues, or modified polypeptides in which one or more residues are modified, and mutants comprising one or more modified residues. Portions or fragments of a FATP can range in size from four amino acid residues to the entire amino acid sequence minus one amino acid and include contiguous portions or  
5 fragments about 4, 5, 6, 7, 8, 9, 10, 15, 25, 30, 40, 50, 75, 100, 150, 200, 300, 400, 500, 600 or more amino acid residues in length. In one particular embodiment, the portion or fragment includes the signature sequence of a FATP polypeptide and is about 360 amino acid residues in length.

The isolated proteins of the invention preferably include mammalian fatty  
10 acid transport proteins of the FATP family of homologous proteins. In one embodiment, the extent of amino acid sequence similarity between a polypeptide having one of the amino acid sequences shown in Figure 45 (SEQ ID NO:47), Figure 47 (SEQ ID NO:49), Figure 112 (SEQ ID NO:117), Figure 51 (SEQ ID NO:53), Figure 53 (SEQ ID NO:55), or Figure 55 (SEQ ID NO:57), and the respective  
15 functional equivalents of these polypeptides is at least about 88%. In other embodiments, the degree of amino acid sequence similarity between a FATP and its respective functional equivalent is at least about 91%, at least about 94%, or at least about 97%.

The polypeptides of the invention also include those FATPs encoded by  
20 polynucleotides which are orthologous to those polynucleotides, the sequences of which are described herein in whole or in part. FATPs which are orthologs to those described herein by amino acid sequence, in whole or in part, are, for example, fatty acid transport proteins 1-6 of dog, rat, chimpanzee, monkey, rabbit, guinea pig, baboon and pig, and are also embodiments of the invention.

25 To determine the percent identity or similarity of two amino acid sequences or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment, and non-homologous (dissimilar) sequences can be disregarded for comparison purposes). In a preferred  
30 embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more

preferably at least 60%, and even more preferably at least 70%, 80%, or 90% of the length of the reference sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or  
5 nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein, amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "similarity"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which  
10 need to be introduced for optimal alignment of the two sequences.

The invention also encompasses polypeptides having a lower degree of identity but having sufficient similarity so as to perform one or more of the same functions performed by the polypeptides described herein by amino acid sequence. Similarity for a polypeptide is determined by conserved amino acid substitution.  
15 Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Conservative substitutions are likely to be phenotypically silent. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues  
20 Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe and Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent is found in Bowie *et al.*, *Science* 247:1306-1310 (1990).

TABLE 1. Conservative Amino Acid Substitutions

5	Aromatic		Phenylalanine	
			Tryptophan	
			Tyrosine	
	Hydrophobic		Leucine	
			Isoleucine	
			Valine	
	Polar		Glutamine	
			Asparagine	
	Basic		Arginine	
			Lysine	
			Histidine	
	Acidic		Aspartic Acid	
			Glutamic Acid	
	Small		Alanine	
			Serine	
			Threonine	
			Methionine	
			Glycine	

The comparison of sequences and determination of percent identity and similarity between two sequences can be accomplished using a mathematical algorithm. (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part 1*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereaux, J., eds., M. Stockton

Press, New York, 1991). In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at <http://www.gcg.com>), using  
5 either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (Devereux, J., *et al.*, *Nucleic Acids Res.* 12(1):387 (1984)) (available at <http://www.gcg.com>), using a  
10 NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Meyers and W. Miller (*CABIOS*, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length  
15 penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.* (*J. Mol.*  
20 *Biol.* 215:403-10 (1990)). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, word length = 12 to obtain nucleotide sequences homologous to (with calculatably significant similarity to) the nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, word length = 3 to obtain amino acid sequences homologous to  
25 the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.*, (*Nucleic Acids Res.* 25(17):3389-3402 (1997)). When utilizing BLAST and gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See <http://www.ncbi.nlm.nih.gov>. Similarity for nucleotide and  
30 amino acid sequences can be defined in terms of the parameters set by the Advanced Blast search available from NCBI (the National Center for Biotechnology

Information; see, for Advanced BLAST page, [www.ncbi.nlm.nih.gov/cgi-bin/BLAST/nph-newblast?Jform=1](http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST/nph-newblast?Jform=1)). These default parameters, recommended for a query molecule of length greater than 85 amino acid residues or nucleotides have been set as follows: gap existence cost, 11, per residue gap cost, 1; lambda ratio,  
5 0.85. Further explanation of version 2.0 of BLAST can be found on related website pages and in Altschul, S.F. *et al.*, *Nucleic Acids Res.* 25:3389-3402 (1997).

In certain embodiments, a contiguous portion can be about 4, 5, 6, 7, 8, 9, 10, 15, 25, 30, 40, 50, 75, 100, 150, 200, 300, 400, 500, 600 or more amino acid residues in length. In one particular embodiment, the portion or fragment includes the  
10 signature sequence of a FATP polypeptide and is about 360 amino acid residues in length.

The invention further relates to fusion proteins, comprising a FATP or functional portion thereof (as described above) as a first moiety, linked to a second moiety not occurring in the FATP as found in nature. Thus, the second moiety can be  
15 an amino acid, peptide or polypeptide. The first moiety can be in an N-terminal location, C-terminal location or internal to the fusion protein. In one embodiment, the fusion protein comprises a FATP as the first moiety, and a second moiety comprising a linker sequence and an affinity ligand. Fusion proteins can be produced by a variety of methods. For example, a fusion protein can be produced by the  
20 insertion of a FATP gene or portion thereof into a suitable expression vector, such as Bluescript SK +/- (Stratagene, La Jolla, CA), pGEX-4T-2 (Pharmacia, Peapack, NJ), pET-24(+) (Novagen, Madison, WI), or vectors of similar construction. The resulting construct can be introduced into a suitable host cell for expression. Upon expression, fusion protein can be purified from cells by means of a suitable affinity matrix (See  
25 e.g., *Current Protocols in Molecular Biology*, Ausubel, F.M. *et al.*, eds., Vol. 2, pp. 16.4.1-16.7.8, containing supplements up through Supplement 42, 1998).

The invention also relates to enzymatically produced, synthetically produced, or recombinantly produced portions of a fatty acid transport protein. Portions of a FATP can be made which have full or partial function on their own, or which when  
30 mixed together (though fully, partially, or nonfunctional alone), spontaneously

assemble with one or more other polypeptides to reconstitute a functional protein having at least one function characteristic of a FATP.

Fragments of a FATP can be produced by direct peptide synthesis, for example those using solid-phase techniques (Roberge, J.Y. *et al.*, *Science* 269:202-204 (1995); Merrifield, J., *J. Am. Chem. Soc.* 85:2149-2154 (1963)). Protein  
5 synthesis can be performed using manual techniques or by automation. Automated synthesis can be carried out using, for instance, an Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Various fragments of a FATP can be synthesized separately and combined using chemical methods.

10 One aspect of the invention is a peptide or polypeptide having the amino acid sequence of a portion of a fatty acid transport protein which is hydrophilic rather than hydrophobic, and ordinarily can be detected as facing the outside of the cell membrane. Such a peptide or polypeptide can be thought of as being an extracellular domain of the FATP, or a mimetic of said extracellular domain. It is known, for  
15 example, that a portion of human FATP4 that includes a highly conserved motif is involved in AMP-CoA binding function (Stuhlsatz-Krouper, S.M. *et al.*, *J. Biol. Chem.* 44:28642-28650 (1998)).

The term "mimetic" as used herein, refers to a molecule, the structure of which is developed from knowledge of the structure of the FATP of interest, or one  
20 or more portions thereof, and, as such, is able to effect some or all of the functions of a FATP.

Portions of a FATP can be prepared by enzymatic cleavage of the isolated protein, or can be made by chemical synthesis methods. Portions of a FATP can also be made by recombinant DNA methods in which restriction fragments, or fragments  
25 that may have undergone further enzymatic processing, or synthetically made DNAs are joined together to construct an altered FATP gene. The gene can be made such that it encodes one or more desired portions of a FATP. These portions of FATP can be entirely homologous to a known FATP, or can be altered in amino acid sequence relative to naturally occurring FATPs to enhance or introduce desired properties such  
30 as solubility, stability, or affinity to a ligand. A further feature of the gene can be a sequence encoding an N-terminal signal peptide directed to the plasma membrane.

An extracellular domain can be determined by a hydrophobicity plot, such as those shown in Figures 28A, 29A, and 35A, or by a hydrophilicity plot such as those shown in Figures 28C, 29C, 35C, 91, 92 and 93. A polypeptide or peptide comprising all or a portion of a FATP extracellular domain can be used in a pharmaceutical composition. When administered to a mammal by an appropriate route, the polypeptide or peptide can bind to fatty acids and compete with the native FATPs in the membrane of cells, thereby making fewer fatty acid molecules available as substrates for transport into cells, and reducing the amount of fatty acids taken up by, for example, the heart, in the case of FATP6.

Another aspect of the invention relates to a method of producing a fatty acid transport protein, variants or portions thereof, and to expression systems and host cells containing a vector appropriate for expression of a fatty acid transport protein.

Cells that express a FATP, a variant or a portion thereof, or an ortholog of a FATP described herein by amino acid sequence, can be made and maintained in culture, under conditions suitable for expression, to produce protein in the cells for cell-based assays, or to produce protein for isolation. These cells can be procaryotic or eucaryotic. Examples of procaryotic cells that can be used for expression include *Escherichia coli*, *Bacillus subtilis* and other bacteria. Examples of eucaryotic cells that can be used for expression include yeasts such as *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pichia pastoris* and other lower eucaryotic cells, and cells of higher eucaryotes such as those from insects and mammals, such as primary cells and cell lines such as CHO, HeLa, 3T3 and BHK cells, preferably COS cells and human kidney 293 cells, and more preferably Jurkat cells. (See, e.g., Ausubel, F.M. *et al.*, eds. *Current Protocols in Molecular Biology*, Greene Publishing Associates and John Wiley & Sons, Inc., containing Supplements up through Supplement 42, 1998)).

In one embodiment, host cells that produce a recombinant FATP, or a portion thereof, a variant, or an ortholog of a FATP described herein by amino acid sequence, can be made as follows. A gene encoding a FATP, variant or a portion thereof can be inserted into a nucleic acid vector, e.g., a DNA vector, such as a plasmid, phage, cosmid, phagemid, virus, virus-derived vector (e.g., SV40, vaccinia, adenovirus, fowl

-41-

pox virus, pseudorabies viruses, retroviruses) or other suitable replicon, which can be present in a single copy or multiple copies, or the gene can be integrated in a host cell chromosome. A suitable replicon or integrated gene can contain all or part of the coding sequence for a FATP or variant, operably linked to one or more expression control regions whereby the coding sequence is under the control of transcription signals and linked to appropriate translation signals to permit translation. The vector can be introduced into cells by a method appropriate to the type of host cells (e.g., transfection, electroporation, infection). For expression from the FATP gene, the host cells can be maintained under appropriate conditions (e.g., in the presence of inducer, normal growth conditions, etc.). Proteins or polypeptides thus produced can be recovered (e.g., from the cells, as in a membrane fraction, from the periplasmic space of bacteria, from culture medium) using suitable techniques. Appropriate membrane targeting signals may be incorporated into the expressed polypeptide. These signals may be endogenous to the polypeptide or they may be heterologous signals.

Polypeptides of the invention can be recovered and purified from cell cultures (or from their primary cell source) by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and high performance liquid chromatography. Known methods for refolding protein can be used to regenerate active conformation if the polypeptide is denatured during isolation or purification.

In a further aspect of the invention are methods for assessing the transport function of any of the fatty acid transport proteins or polypeptides described herein, including orthologs, and in variations of these, methods for identifying an inhibitor (or an enhancer) of such function and methods for assessing the transport function in the presence of a candidate inhibitor or a known inhibitor.

A variety of systems comprising living cells can be used for these methods. Cells to be used in fatty acid transport assays, and further in methods for identifying an inhibitor or enhancer of this function, express one or more FATPs. See Examples



3, 6, 9, 12 and 14 for data on tissue distribution of expression of FATPs, and Examples 10 and 11 describing recombinant cells expressing FATP. Cells for use in cell-based assays described herein can be drawn from a variety of sources, such as isolated primary cells of various organs and tissues wherein one or more FATPs are naturally expressed. In some cases, the cells can be from adult organs, and in some cases, from embryonic or fetal organs, such as heart, lung, liver, intestine, skeletal muscle, kidney and the like. Cells for this purpose can also include cells cultured as fragments of organs or in conditions simulating the cell type and/or tissue organization of organs, in which artificial materials may be used as substrates for cell growth. Other types of cells suitable for this purpose include cells of a cell strain or cell line (ordinarily comprising cells considered to be "transformed") transfected to express one or more FATPs.

A further embodiment of the invention is a method for detecting, in a sample of cells, a fatty acid transport protein, a portion or fragment thereof, a fusion protein comprising a FATP or a portion thereof, or an ortholog as described herein, wherein the cells can be, for instance, cells of a tissue, primary culture cells, or cells of a cell line, including cells into which nucleic acid has been introduced. The method comprises adding to the sample an agent that specifically binds to the protein, and detecting the agent specifically bound to the protein. Appropriate washing steps can be added to reduce nonspecific binding to the agent. The agent can be, for example, an antibody, a ligand or a substrate mimic. The agent can have incorporated into it, or have bound to it, covalently or by high affinity non-covalent interactions, for instance, a label that facilitates detection of the agent to which it is bound, wherein the label can be, but is not limited to, a phosphorescent label, a fluorescent label, a biotin or avidin label, or a radioactive label. The means of detection of a fatty acid transport protein can vary, as appropriate to the agent and label used. For example, for an antibody that binds to the fatty acid transport protein, the means of detection may call for binding a second antibody, which has been conjugated to an enzyme, to the antibody which binds the fatty acid transport protein, and detecting the presence of the second antibody by means of the enzymatic activity of the conjugated enzyme.

Similar principles can also be applied to a cell lysate or a more purified preparation of proteins from cells that may comprise a fatty acid transport protein of interest, for example in the methods of immunoprecipitation, immunoblotting, immunoaffinity methods, that in addition to detection of the particular FATP, can also be used in purification steps, and qualitative and quantitative immunoassays. See, for instance, chapters 11 through 14 in *Antibodies: A Laboratory Manual*, E. Harlow and D. Lane, eds., Cold Spring Harbor Laboratory, 1988.

Isolated fatty acid transport protein or, an antigenically similar portion thereof, especially a portion that is soluble, can be used in a method to select and identify molecules which bind specifically to the FATP. Fusion proteins comprising all of, or a portion of, the fatty acid transport protein linked to a second moiety not occurring in the FATP as found in nature, can be prepared for use in another embodiment of the method. Suitable fusion proteins for this purpose include those in which the second moiety comprises an affinity ligand (e.g., an enzyme, antigen, epitope). FATP fusion proteins can be produced by the insertion of a gene encoding the FATP or a variant thereof, or a suitable portion of such gene into a suitable expression vector, which encodes an affinity ligand (e.g., pGEX-4T-2 and pET-15b, encoding glutathione S-transferase and His-Tag affinity ligands, respectively). The expression vector can be introduced into a suitable host cell for expression. Host cells are lysed and the lysate, containing fusion protein, can be bound to a suitable affinity matrix by contacting the lysate with an affinity matrix. In a particular embodiment, a nucleic acid encodes a portion of a FATP polypeptide which includes a motif or domain, for example, a lipocalin domain or an AMP-binding domain. Such a polypeptide portion can be a functional portion of a FATP protein. The term "lipocalin domain" is an art recognized term and as used herein refers to a particular domain present in FATP proteins. This domain is described as including regions of sequence homology as well as a common tertiary structure represented as an eight stranded antiparallel beta-barrel. (see Banaszak, L. *et al.*, *Advances in Protein Chemistry*, 45: 89-151). Many lipocalin domains can be identified structurally as a sequence contained within the general formula: [DENG]-X-[DENQGSTARK]-X(0,2)-[DENQARK]-[LIVFY]-{CP}-G-{C}-W-[FYWLRH-X]-[LIVMTA], e.g., the lipocalin signature sequence or

-44-

consensus pattern (SEQ ID NO: 125). One skilled in the art will recognize that a lipocalin domain for a particular FATP protein can vary in sequence from this general formula. A FATP lipocalin domain can be, for example, identical to the lipocalin signature sequence, or, can exhibit 60, 65, 70, 75, 80, 85, 90, 95 or greater sequence  
5 percent identity in comparison to the general formula provided that it still retains the necessary lipocalin binding function.

For example, a lipocalin domain for each of the human FATPs, hsFATP1 (SEQ ID NO: 126), hsFATP2 (SEQ ID NO: 127), hsFATP3 (SEQ ID NO: 128), hsFATP4 (SEQ ID NO: 129), hsFATP5 (SEQ ID NO: 130), and hsFATP6 (SEQ ID  
10 NO: 131) has been identified. These particular lipocalin domains are located near the N-terminal portion of the specified proteins (see Figure 118). The sequences of these lipocalin domains are highly conserved across the FATP family. A search using the lipocalin signature sequence conducted on a public database ([www.ebi.ac.uk/interpro/](http://www.ebi.ac.uk/interpro/)), indicated that the lipocalin domains of hsFATP1 and  
15 hsFATP4 share identity with signature sequence. In addition, a search directed to identifying sequences having at least 80% identity to the lipocalin signature sequence identified three additional human FATPs, hsFATP3, hsFATP5 and hsFATP6.

A lipocalin domain can also be identified functionally since, for example, it has been identified as a binding motif capable of binding fatty acids. In particular,  
20 the studies described in Experiment 20 demonstrated that fusion proteins including the lipocalin domains from hsFATP4 bound long chain fatty acids such as oleates and palmitates with great specificity. Other fatty acids can also be used to assess binding in FATP4 and other members of the FATP family.

Polypeptides, including fusion polypeptides, which contain a lipocalin domain  
25 can also include additional components. For example, fusion polypeptides containing a lipocalin domain can include amino acid residues from the portion of the protein which is located upstream, *i. e.*, in the direction of the N-terminal end of a FATP protein, from the lipocalin domain. As the term "upstream sequences" is used herein in relation to the lipocalin domain, it is intended to refer to the amino acid residues of  
30 a FATP protein which are located between the signal peptide (when one is present) and the lipocalin domain. In the absence of a signal peptide, the term refers to the

portion of a FATP protein between the lipocalin domain and the amino terminus (see Figure 115).

Fusion polypeptides which contain a lipocalin domain can also include additional domains or motifs, for example, an AMP binding domain can be included.

- 5 For example, an AMP binding domain for each of the human FATPs, hsFATP1 (SEQ ID NO: 132), hsFATP2 (SEQ ID NO: 133), hsFATP3 (SEQ ID NO: 134), hsFATP4 (SEQ ID NO: 135), hsFATP5 (SEQ ID NO: 136) and hsFATP6 (SEQ ID NO: 137) has been identified (see Figure 118).

- In one embodiment, the fusion protein can be immobilized on a suitable  
10 affinity matrix under conditions sufficient to bind the affinity ligand portion of the fusion protein to the matrix, and is contacted with one or more candidate binding agents (e.g., a mixture of peptides) to be tested, under conditions suitable for binding of the binding agents to the FATP portion of the bound fusion protein. Next, the affinity matrix with bound fusion protein can be washed with a suitable wash buffer  
15 to remove unbound candidate binding agents and non-specifically bound candidate binding agents. Those agents which remain bound can be released by contacting the affinity matrix with fusion protein bound thereto with a suitable elution buffer. Wash buffer can be formulated to permit binding of the fusion protein to the affinity matrix, without significantly disrupting binding of specifically bound binding agents. In this  
20 aspect, elution buffer can be formulated to permit retention of the fusion protein by the affinity matrix, but can be formulated to interfere with binding of the candidate binding agents to the target portion of the fusion protein. For example, a change in the ionic strength or pH of the elution buffer can lead to release of specifically bound agent, or the elution buffer can comprise a release component or components  
25 designed to disrupt binding of specifically bound agent to the target portion of the fusion protein.

- Immobilization can be performed prior to, simultaneous with, or after, contacting the fusion protein with candidate binding agent, as appropriate. Various permutations of the method are possible, depending upon factors such as the  
30 candidate molecules tested, the affinity matrix-ligand pair selected, and elution buffer formulation. For example, after the wash step, fusion protein with binding agent

-46-

molecules bound thereto can be eluted from the affinity matrix with a suitable elution buffer (a matrix elution buffer, such as glutathione for a GST fusion). Where the fusion protein comprises a cleavable linker, such as a thrombin cleavage site, cleavage from the affinity ligand can release a portion of the fusion with the candidate agent bound thereto. Bound agent molecules can then be released from the fusion protein or its cleavage product by an appropriate method, such as extraction.

One or more candidate binding agents can be tested simultaneously. Where a mixture of candidate binding agents is tested, those found to bind by the foregoing processes can be separated (as appropriate) and identified by suitable methods (e.g., PCR, sequencing, chromatography). Large libraries of candidate binding agents (e.g., peptides, RNA oligonucleotides) produced by combinatorial chemical synthesis or by other methods can be tested (see e.g., Ohlmeyer, M.H.J. *et al.*, *Proc. Natl. Acad. Sci. USA* 90:10922-10926 (1993) and DeWitt, S.H. *et al.*, *Proc. Natl. Acad. Sci. USA* 90:6909-6913 (1993), relating to tagged compounds; see also Rutter, W.J. *et al.* U.S. Patent No. 5,010,175; Huebner, V.D. *et al.*, U.S. Patent No. 5,182,366; and Geysen, H.M., U.S. Patent No. 4,833,092). Random sequence RNA libraries (see Ellington, A.D. *et al.*, *Nature* 346:818-822 (1990); Bock, L.C. *et al.*, *Nature* 355:584-566 (1992); and Szostak, J.W., *Trends in Biochem. Sci.* 17:89-93 (March, 1992)) can also be screened according to the present method to select RNA molecules which bind to a target FATP or FATP fusion protein. Where binding agents selected from a combinatorial library by the present method carry unique tags, identification of individual biomolecules by chromatographic methods is possible. Where binding agents do not carry tags, chromatographic separation, followed by mass spectrometry to ascertain structure, can be used to identify binding agents selected by the method, for example.

The invention also comprises a method for identifying an agent which inhibits interaction between a fatty acid transport protein (e.g., one comprising the amino acid sequence in SEQ ID NO:47, SEQ ID NO:49, SEQ ID NO:117, SEQ ID NO:53, SEQ ID NO:55, or SEQ ID NO:57), and a ligand of said protein. The FATP can be one described by an amino acid sequence herein, a portion or fragment thereof, a variant thereof, or an ortholog thereof, or a FATP fusion protein. Here, a ligand can be, for

instance, a substrate, or a substrate mimic, an antibody, or a compound, such as a peptide, that binds with specificity to a site on the protein. The method comprises combining, not limited to a particular order, the fatty acid protein, the ligand of the protein, and a candidate agent to be assessed for its ability to inhibit interaction  
5 between the protein and the ligand, under conditions appropriate for interaction between the protein and the ligand (e.g., pH, salt, temperature conditions conducive to appropriate conformation and molecular interactions); determining the extent to which the protein and ligand interact; and comparing (1) the extent of protein-ligand interaction in the presence of candidate agent with (2) the extent of protein-ligand  
10 interaction in the absence of candidate agent, wherein if (1) is less than (2), then the candidate agent is one which inhibits interaction between the protein and the ligand.

The method can be facilitated, for example, by using an experimental system which employs a solid support (column chromatography matrix, wall of a plate, microtiter wells, column pore glass, pins to be submerged in a solution, beads, etc.)  
15 to which the protein can be attached. Accordingly, in one embodiment, the protein can be fixed to a solid phase directly or indirectly, by a linker. The candidate agent to be tested is added under conditions conducive for interaction and binding to the protein. The ligand is added to the solid phase system under conditions appropriate for binding. Excess ligand is removed, as by a series of washes done under  
20 conditions that do not disrupt protein-ligand interactions. Detection of bound ligand can be facilitated by using a ligand that carries a label (e.g., fluorescent, chemiluminescent, radioactive). In a control experiment, protein and ligand are allowed to interact in the absence of any candidate agent, under conditions otherwise identical to those used for the "test" conditions where candidate inhibiting agent is  
25 present, and any washes used in the test conditions are also used in the control. The extent to which ligand binds to the protein in the presence of candidate agent is compared to the extent to which ligand binds to the protein in the absence of the candidate agent. If the extent to which interaction of the protein and the ligand occurs is less in the presence of the candidate agent than in the absence of the  
30 candidate agent, the candidate agent is an agent which inhibits interaction between the protein and the ligand of the protein.

In a further embodiment, an inhibitor (or an enhancer) of a fatty acid transport protein can be identified. The method comprises steps which are, or are variations of the following: contacting the cells with fatty acid, wherein the fatty acid can be labeled for convenience of detection; contacting a first aliquot of the cells with an agent being tested as an inhibitor (or enhancer) of fatty acid uptake while maintaining a second aliquot of cells under the same conditions but without contact with the agent; and measuring (e.g., quantitating) fatty acid in the first and second aliquots of cells; wherein a lesser quantity of fatty acid in the first aliquot compared to that in the second aliquot is indicative that the agent is an inhibitor of fatty acid uptake by a fatty acid transport protein. A greater quantity of fatty acid in the first aliquot compared to that in the second aliquot is indicative that the agent is an enhancer of fatty acid uptake by a fatty acid transport protein.

A particular embodiment of identifying an inhibitor or enhancer of fatty acid transport function employs the above steps, but also employs additional steps preceding those given above: introducing into cells of a cell strain or cell line ("host cells" for the intended introduction of, or after the introduction of, a vector) a vector comprising a fatty acid transport protein gene, wherein expression of the gene can be regulatable or constitutive, and providing conditions to the host cells under which expression of the gene can occur.

The terms "contacting" and "combining" as used herein in the context of bringing molecules into close proximity to each other, can be accomplished by conventional means. For example, when referring to molecules that are soluble, contacting is achieved by adding the molecules together in a solution. "Contacting" can also be adding an agent to a test system, such as a vessel containing cells in tissue culture.

The term "inhibitor" or "antagonist", as used herein, refers to an agent which blocks, diminishes, inhibits, hinders, limits, decreases, reduces, restricts or interferes with fatty acid transport into the cytoplasm of a cell, or alternatively and additionally, prevents or impedes the cellular effects associated with fatty acid transport. The term "enhancer" or "agonist", as used herein, refers to an agent which augments, enhances, or increases fatty acid transport into the cytoplasm of a cell. An antagonist will

decrease fatty acid concentration, fatty acid metabolism and byproduct levels in the cell, leading to phenotypic and molecular changes.

In order to produce a "host cell" type suitable for fatty acid uptake assays and for assays derived therefrom for identifying inhibitors or enhancers thereof, a nucleic acid vector can be constructed to comprise a gene encoding a fatty acid transport protein, for example, human FATP1, FATP2, FATP3, FATP4, FATP5, FATP6, a mutant or variant thereof, an ortholog of the human proteins, such as mouse orthologs or orthologs found in other mammals, or a FATP family protein of origin in an organism other than a mammal. The gene of the vector can be regulatable, such as by the placement of the gene under the control of an inducible or repressible promoter in the vector (e.g., inducible or repressible by a change in growth conditions of the host cell harboring the vector, such as addition of inducer, binding or functional removal of repressor from the cell milieu, or change in temperature) such that expression of the FATP gene can be turned on or initiated by causing a change in growth conditions, thereby causing the protein encoded by the gene to be produced, in host cells comprising the vector, as a plasma membrane protein. Alternatively, the FATP gene can be constitutively expressed.

A vector comprising a FATP gene, such as a vector described herein, can be introduced into host cells by a means appropriate to the vector and to the host cell type. For example, commonly used methods such as electroporation, transfection, for instance, transfection using  $\text{CaCl}_2$ , and transduction (as for a virus or bacteriophage) can be used. Host cells can be, for example, mammalian cells such as primary culture cells or cells of cell lines such as COS cells, 293 cells or Jurkat cells. Host cells can also be, in some cases, cells derived from insects, cells of insect cell lines, bacterial cells, such as *E. coli*, or yeast cells, such as *S. cerevisiae*. It is preferred that the fatty acid transport protein whose function is to be assessed, with or without a candidate inhibitor or enhancer, be produced in host cells whose ancestor cells originated in a species related to the species of origin of the FATP gene encoding the fatty acid transport protein. For example, it is preferable that tests of function or of inhibition or enhancement of a mammalian FATP be carried out in host mammalian cells producing the FATP, rather than bacterial cells or yeast cells.



Host cells comprising a vector comprising a regulatable FATP gene can be treated so as to allow expression of the FATP gene and production of the encoded protein (e.g., by contacting the cells with an inducer compound that effects transcription from an inducible promoter operably linked to the FATP gene).

5           Alternatively, host cells containing an endogenous FATP gene can be engineered to activate or deactivate expression of the FATP gene and production of the encoded protein. For example, homologous recombination, often referred to as targeting, can be utilized to alter the regulatory region associated with the FATP gene to increase or decrease the level of expression. Alteration of the regulatory region  
10       can include disablement of the regulatory region associated with the FATP gene and/or replacement of the region or a portion of the region. A variety of regulatory regions are known which can be transfected into cells to cause an endogenous gene to display a pattern of induction or expression that differs from that of the cell prior to transfection.

15           The test agent (e.g., an agonist or antagonist) is added to the cells to be used in a fatty acid transport assay, in the presence or absence of test agent, under conditions suitable for production and/or maintenance of the expressed FATP in a conformation appropriate for association of the FATP with test agent and substrate. For example, conditions under which an agent is assessed, such as media and  
20       temperature requirements, can, initially, be similar to those necessary for transport of typical fatty acid substrates across the plasma membrane. One of ordinary skill in the art will know how to vary experimental conditions depending upon the biochemical nature of the test agent. The test agent can be added to the cells in the presence of fatty acid, or in the absence of fatty acid substrate, with the fatty acid substrate being  
25       added following the addition of the test agent. The concentration at which the test agent can be evaluated can be varied, as appropriate, to test for an increased effect with increasing concentrations.

Test agents to be assessed for their effects on fatty acid transport can be any chemical (element, molecule, compound), made synthetically, made by recombinant  
30       techniques or isolated from a natural source. For example, test agents can be peptides, polypeptides, peptoids, sugars, hormones, or nucleic acid molecules, such as

antisense nucleic acid molecules. In addition, test agents can be small molecules or molecules of greater complexity made by combinatorial chemistry, for example, and compiled into libraries. These libraries can comprise, for example, alcohols, alkyl halides, amines, amides, esters, aldehydes, ethers and other classes of organic compounds. Test agents can also be natural or genetically engineered products isolated from lysates of cells, bacterial, animal or plant, or can be the cell lysates themselves. Presentation of test compounds to the test system can be in either an isolated form or as mixtures of compounds, especially in initial screening steps.

Thus, the invention relates to a method for identifying agents which alter fatty acid transport, the method comprising providing the test agent to the cell (wherein "cell" includes the plural, and can include cells of a cell strain, cell line or culture of primary cells or organ culture, for example), under conditions suitable for binding to its target, whether to the FATP itself or to another target on or in the cell, wherein the transformed cell comprises a FATP.

In greater detail, to test one or more agents or compounds (e.g., a mixture of compounds can conveniently be screened initially) for inhibition of the transport function of a fatty acid transport protein, the agent(s) can be contacted with the cells. The cells can be contacted with a labeled fatty acid. The fatty acid can be, for example, a known substrate of the fatty acid transport protein such as oleate or palmitate. The fatty acid can itself be labeled with a radioactive isotope, (e.g.,  $^3\text{H}$  or  $^{14}\text{C}$ ) or can have a radioactively labeled adduct attached. In other variations, the fatty acid can have chemically attached to it a fluorescent label, or a substrate for an enzyme occurring within the cells, wherein the substrate yields a detectable product, such as a highly colored or fluorescent product. Addition of candidate inhibitors and labeled substrate to the cells comprising fatty acid transport protein can be in either order or can be simultaneous.

A second aliquot of cells, which can be called "control" cells (a "first" aliquot of cells can be called "test" cells), is treated, if necessary (as in the case of transformed "host" cells), so as to allow expression of the FATP gene, and is contacted with the labeled substrate of the fatty acid transport protein. The second aliquot of cells is not contacted with one or more agents to be tested for inhibition of

-52-

the transport function of the protein produced in the cells, but is otherwise kept under the same culture conditions as the first aliquot of cells.

In a further step of a method to identify inhibitors of a fatty acid transport protein, the labeled fatty acid is measured in the first and second aliquots of cells. A preliminary step of this measurement process can be to separate the external medium from the cells so as to be able to distinguish the labeled fatty acid external to the cells from that which has been transported inside the cells. This can be accomplished, for instance, by removing the cells from their growth container, centrifuging the cell suspension, removing the supernatant and performing one or more wash steps to extensively dilute the remaining medium which may contain labeled fatty acid. Detection of the labeled fatty acid can be by a means appropriate to the label used. For example, for a radioactive label, detection can be by scintillation counting of appropriately prepared samples of cells (e.g., lysates or protein extracts); for a fluorescent label, by measuring fluorescence in the cells by appropriate instrumentation.

If a compound tested as a candidate inhibitor of transport function causes the test cells to have less labeled fatty acid detected in the cells than that detected in the control cells, then the compound is an inhibitor of the fatty acid transport protein. Procedures analogous to those above can be devised for identifying enhancers (agonists of FATPs) of fatty acid transport function wherein if the test cells contain more labeled fatty acid than that detected in the control cells, or if the fatty acid is taken up at a higher rate, then the compound being tested can be concluded to be an enhancer of the fatty acid transport protein.

Example 13 describes use of an assay of this type to identify an inhibitor of a FATP. In Example 13, an antisense oligonucleotide which specifically inhibits biosynthesis of mmFATP4 was demonstrated to inhibit fatty acid uptake into mouse enterocytes. Similarly, antisense oligonucleotides directed towards specifically inhibiting the biosynthesis of FATP6 in heart cells, FATP5 in liver cells, FATP3 in lung cells, and FATP2 in colon cells, can be demonstrated as examples of "test agents" that inhibit fatty acid transport.

Another assay to determine whether an agent is an inhibitor (or enhancer) of fatty acid transport employs animals, one or more of which are administered the agent, and one or more of which are maintained under similar conditions, but are not administered the agent. Both groups of animals are given fatty acids (e.g., orally,  
5 intravenously, by tube inserted into stomach or intestine), and the fatty acids taken up into a bodily fluid (e.g., serum) or into an organ or tissue of interest are measured from comparable samples taken from each group of animals. The fatty acids may carry a label (e.g., radioactive) to facilitate detection and quantitation of fatty acids taken up into the fluid or tissue being sampled. This type of assay can be used alone  
10 or can be used in addition to *in vitro* assays of a candidate inhibitor or enhancer.

An agent determined to be an inhibitor (or enhancer) of FATP function, such as fatty acid binding and/or fatty acid uptake, can be administered to cells in culture, or *in vivo*, to a mammal (e.g. human) to inhibit (or enhance) FATP function. Such an agent may be one that acts directly on the FATP (for example, by  
15 binding) or can act on an intermediate in a biosynthetic pathway to produce FATP, such as transcription of the FATP gene, processing of the mRNA, or translation of the mRNA. An example of such an agent is antisense oligonucleotide.

Antisense methods similar to those illustrated in Example 13 can be used to determine the target FATP of a compound or agent that has an inhibitory or  
20 enhancing effect on fatty acid uptake. For example, antisense oligonucleotide directed to the inhibition of FATP4 biosynthesis can be added to lung cells or cell lines derived from lung cells. In addition, antisense oligonucleotides directed to the inhibition of other FATPs, except for FATP3, can also be added to the lung cells. The administration of antisense oligonucleotides in this manner ensures that the  
25 predominant FATP activity remaining in the cells comes from FATP3. After a period of incubation of the cells with the antisense oligonucleotides sufficient to deplete the plasma membrane of the FATPs whose biosynthesis has been inhibited, a test agent, preferably one that has been shown by some preliminary test to have an inhibitory or enhancing activity on fatty acid transport, can be added to the lung cells. If the test  
30 agent is now demonstrated, after treatment of the cells with antisense oligonucleotides, to have an inhibitory or enhancing activity on fatty acid transport in

the lung cells, it can be concluded that the target of the test agent is FATP3, or a molecule involved in the biosynthesis or activity of FATP3.

In another type of cell-based assay for uptake of fatty acids, a change of intracellular pH resulting from the uptake of fatty acids can be followed by an indicator fluorophore. The fluorophore can be taken up by the cells in a preincubation step. Fatty acids can be added to the cell medium, and after some period of incubation to allow FATP-mediated uptake of fatty acids, the change in  $\lambda_{\max}$  of fluorescence can be measured, as an indicator of a change in intracellular pH, as the  $\lambda_{\max}$  of fluorescence of the fluorophore changes with the pH of its environment, thereby indicating uptake of fatty acids. One such fluorophore is BCECF (2', 7'-bis(2-carboxyethyl)-5(6)-carboxyfluorescein; Rink, T.J. *et al.*, *J. Cell. Biol.* 95: 189 (1982)).

In assays similar to those described above, a candidate inhibitor or enhancer of fatty acid transport function can be added (or mock-added, for control cultures) to cultures of cells engineered to express a desired FATP to which fatty acid substrate is also added. Inhibition of fatty acid uptake is indicated by a lack of the drop in pH, indicating fatty acid uptake, that is seen in control cells. Enhancement of fatty acid uptake is indicated by a decrease in intracellular pH, as compared to control cells not receiving the candidate enhancer of fatty acid transport function.

Yeast cells can be used in a similar cell-based assay for the uptake of fatty acids mediated by a FATP, and such an assay can be adapted to a screening assay for the identification of agents that inhibit or enhance fatty acid uptake by an FATP. Yeast cells lacking an endogenous FATP activity (mutated, disrupted or deleted for *FAT1*; Faergeman, N.J. *et al.*, *J. Biol. Chem.* 272(13):8531-8538 (1997); Watkins, P.A. *et al.*, *J. Biol. Chem.* 273(29):18210-18219 (1998)) can be engineered to harbor a related gene of the family of FATP-encoding genes, such as a mammalian FATP (e.g., human FATP4).

Examples of expression vectors include pEG (Mitchell, D.A., *et al.*, *Yeast* 9:715-723 (1993)) and pDAD1 and pDAD2, which contain a *GALI* promoter (Davis, L. I. and Fink, G. R., *Cell* 61:965-978 (1990)). A variety of promoters are suitable for expression. Available yeast vectors offer a choice of promoters. In one

embodiment, the inducible *GAL1* promoter is used. In another embodiment, the constitutive *ADHI* promoter (alcohol dehydrogenase; Bennetzen, J. L. and Hall, B. D., *J. Biol. Chem.* 257:3026-3031 (1982)) can be used to express an inserted gene on glucose-containing media. An example of a vector suitable for expression of a  
5 heterologous FATP gene in yeast is pQB169.

With the introduced FATP gene providing the only fatty acid transport protein function for the yeast cells, it is possible to study effect of the heterologous FATP on fatty acid transport into the yeast cells in isolation. Assays for the uptake of fatty acids into the yeast cells can be devised that are similar to those described above  
10 and/or those assays that have been illustrated in the Examples. Tests for candidate inhibitors or enhancers of the heterologous FATP can be done in cultures of yeast cells, wherein the yeast cells are incubated with fatty acid substrate and an agent to be tested as an inhibitor or enhancer of FATP function. FATP uptake after a period of time can be measured by analyzing the contents of the yeast cells for fatty acid  
15 substrate, as compared with control yeast cells incubated with the fatty acid, but not with the test agent. Yeast cells have the additional advantage, over mammalian cells in culture, for example, that yeast cells can be forced to rely upon fatty acids as their only source of carbon, if the growth medium supplied to the yeast cells is formulated to contain no other source of carbon. Thus, the effect of the heterologous FATP on  
20 fatty acid uptake and metabolism in the engineered yeast cells can be amplified. An agent that efficiently blocks transport function of the heterologous FATP could result in death of the yeast cells. Thus, in this case, inhibition of function of the heterologous FATP can result in loss of viability. A simple measure of viability is turbidity of the yeast suspension culture, which can be adapted to a high throughput  
25 screening assay for effects of various agents to be tested, using microtiter plates or similar devices for small-volume cultures of the engineered yeast cells.

Cell-free assays can also be used to measure the transport of fatty acids across a membrane, and therefor also to assess a test treatment or test agent for its effect on the rate or extent of fatty acid transport. An isolated FATP, for example in the  
30 presence of a detergent that preserves the native 3-dimensional structure of the FATP, or partially purified FATP, can be used in an artificial membrane system typically

-56-

used to preserve the native conformation and activity of membrane proteins. Such systems include liposomes, artificial bilayers of phospholipids, isolated plasma membrane such as cell membrane fragments, cell membrane fractions, or cell membrane vesicles, and other systems in which the FATP can be properly oriented within the membrane to have transport activity. Assays for transport activity can be performed using methods analogous to those that can be used in cells engineered to predominantly express one FATP whose function is to be measured. A labeled (e.g., radioactively labeled) fatty acid substrate can be incubated with one side of a bilayer or in a suspension of liposomes constructed to integrate a properly oriented FATP.

5    The accumulation of fatty acids with time can be measured, using appropriate means to detect the label (e.g., scintillation counting of medium on each side of the bilayer, or of the contents of liposomes isolated from the surrounding medium). Assays such as these can be adapted to use for the testing of agents which might interact with the FATP to produce an inhibitory or an enhancing effect on the rate or extent of fatty acid transport. That is, the above-described assay can be done in the presence or

10    absence of the agent to be tested, and the results compared.

For examples of isolation of membrane proteins (ADP/ATP carrier and uncoupling protein), reconstitution into phospholipid vesicles, and assays of transport, see Klingenberg, M. *et al.*, *Methods Enzymol.* 260:369-389 (1995). For an example of a membrane protein (phosphate carrier of *Saccharomyces cerevisiae*) that was purified and solubilized from *E. coli* inclusion bodies, see Schroer, A. *et al.*, *J. Biol. Chem.* 273: 14269-14276 (1998). The Glut1 glucose transporter of rat has been expressed in yeast. A crude membrane fraction of the yeast was prepared and reconstituted with soybean phospholipids into liposomes. Glucose transport activity

20    could be measured in the liposomes (Kasahara, T. and Kasahara, M., *J. Biol. Chem.* 273: 29113-29117 (1998)). Similar methods can be applied to the proteins and polypeptides of the invention.

Another embodiment of the invention is a method for inhibiting fatty acid uptake in a mammal (e.g., a human), comprising administering to the mammal a therapeutically effective amount of an inhibitor of the transport function of one or

30    more of the fatty acid transport proteins, thereby decreasing fatty acid uptake by cells

comprising the fatty acid protein(s). Where it is desirable to reduce the uptake of fatty acids, for example, in the treatment of chronic obesity or as a part of a program of weight control or hyperlipidemia control in a human, one or more inhibitors of one or more of the fatty acid transport proteins can be administered in an effective dose, and by an effective route, for example, orally, or by an indwelling device that can deliver doses to the small intestine. The inhibitor can be one identified by methods described herein, or can be one that is, for instance, structurally related to an inhibitor identified by methods described herein (e.g., having chemical adducts to better stabilize or solubilize the inhibitor). The invention further relates to compositions comprising inhibitors of fatty acid uptake in a mammal, which may further comprise pharmaceutical carriers suitable for administration to a subject mammal, such as sterile solubilizing or emulsifying agents.

A further embodiment of the present invention is a method of enhancing or increasing fatty acid uptake, such as enhancing or increasing LCFA uptake in the small intestine (e.g., to treat or prevent a malabsorption syndrome or other wasting condition) or in the liver (e.g., by an enhancer of FATP5 transport activity to treat acute liver failure) or in the kidney (e.g., by an enhancer of FATP2 transport activity to treat kidney failure). In this embodiment, a therapeutically effective amount of an enhancer of the transport function of one or more of the fatty acid transport proteins can be administered to a mammalian subject, with the result that fatty acid uptake in the small intestine is enhanced. In this embodiment, one or more enhancers of one or more of fatty acid transport proteins is administered in an effective dose and by a route (e.g., orally or by a device, such as an indwelling catheter or other device) which can deliver doses to the gut. The enhancer of FATP function (e.g., an enhancer of FATP4 function) can be identified by methods described herein or can be one that is structurally similar to an enhancer identified by methods described herein.

Aerobic reperfusion of ischemic myocardium is a common clinical event which can occur during such treatments as cardiac surgery, angioplasty, and thrombolytic therapy after a myocardial infarction. During reperfusion, a rapid recovery of myocardial energy production is essential for the complete recovery of contractile function. Not only the extent of recovery of myocardial energy



metabolism but also the type of energy substrate used by the heart during reperfusion are important determinants of functional recovery. Circulating fatty acid levels increase following acute myocardial infarction or during cardiac surgery, such that during and following ischemia the heart muscle can be exposed to very high concentrations of fatty acids (Lopaschuk, G.D. and W. C. Stanley, *Science and Medicine* (November/December 1997)). High plasma fatty acid concentrations increase the severity of ischemic damage in a number of experimental models of cardiac ischemia and have been linked to depression of mechanical function during aerobic reperfusion of previously ischemic hearts. Further data show that modifying fatty acid utilization can be beneficial for heart function in ischemia and can be a useful approach for the treatment of angina. See, e.g., Desideri and Celegon, *Am. J. Cardiol.* 82(5A):50K-53K; Lopaschuk, *Am. J. Cardiol.* 82(5A):14K-17K. Plasma fatty acid concentrations can be reduced by administering to a human subject or other mammal an effective amount of an inhibitor of a FATP such as FATP2 or FATP4, thereby providing a way of reducing fatty acid utilization by the heart.

In a further embodiment of the invention, a therapeutically effective amount of an inhibitor of hsFATP6 can be administered to a human patient by a suitable route, to reduce the uptake of fatty acids by cardiac muscle. This treatment is desirable in patients who are diagnosed as having, or who are at risk of, abnormal accumulations of fatty acids in the heart or a detrimentally high rate of uptake of fatty acids into the heart, because of ischemic heart disease, or following ischemia or trauma to the heart.

The invention further relates to antibodies that bind to an isolated or recombinant fatty acid transport protein of the FATP family, including portions of antibodies, which can specifically recognize and bind to one or more FATPs. The antibodies and portions thereof of the invention include those which bind to one or more FATPs of mouse or other mammalian species. In a preferred embodiment, the antibodies specifically bind to a naturally occurring FATP of humans. The antibodies can be used in methods to detect or to purify a protein of the present invention or a portion thereof by various methods of immunoaffinity chromatography, to inhibit the

function of a protein in a method of therapy, or to selectively inactivate an active site, or to study other aspects of the structure of these proteins, for example.

The antibodies of the present invention can be polyclonal or monoclonal. The term antibody is intended to encompass both polyclonal and monoclonal antibodies.

5 Antibodies of the present invention can be raised against an appropriate immunogen, including proteins or polypeptides of the present invention, such as an isolated or recombinant FATP1, FATP2, FATP3, FATP4, FATP5, FATP6, mtFATP, ceFATPa, ceFATPb, scFATP or portions thereof, or synthetic molecules, such as synthetic peptides (e.g., conjugated to a suitable carrier). Preferred embodiments are antibodies  
10 that bind to any of the following: hsFATP1, hsFATP2, hsFATP3, hsFATP4, hsFATP5 or hsFATP6. The immunogen can be a polypeptide comprising a portion of a FATP and having at least one function of a fatty acid transport protein, as described herein.

The term antibody is also intended to encompass single chain antibodies,  
15 chimeric, humanized or primatized (CDR-grafted) antibodies and the like, as well as chimeric or CDR-grafted single chain antibodies, comprising portions from more than one species. For example, the chimeric antibodies can comprise portions of proteins derived from two different species, joined together chemically by conventional techniques or prepared as a single contiguous protein using genetic  
20 engineering techniques (e.g., DNA encoding the protein portions of the chimeric antibody can be expressed to produce a contiguous protein chain. See, e.g., Cabilly *et al.*, U.S. Patent No. 4,816,567; Cabilly *et al.*, European Patent No. 0,125,023 B1; Boss *et al.*, U.S. Patent No. 4,816,397; Boss *et al.*, European Patent No. 0,120,694 B1; Neuberger, M.S. *et al.*, WO 86/01533; Neuberger, M.S. *et al.*, European Patent  
25 No. 0,194,276 B1; Winter, U.S. Patent No. 5,225,539; Winter, European Patent No. 0,239,400 B1; Queen *et al.*, U.S. Patent No. 5,585,089; and Queen *et al.*, European Patent No. EP 0 451 216 B1. See also, Newman, R. *et al.*, *BioTechnology*, 10:1455-1460 (1992), regarding primatized antibody, and Ladner *et al.*, U.S. Patent No. 4,946,778 and Bird, R.E. *et al.*, *Science*, 242:423-426 (1988) regarding single chain  
30 antibodies.)

-60-

Whole antibodies and biologically functional fragments thereof are also encompassed by the term antibody. Biologically functional antibody fragments which can be used include those fragments sufficient for binding of the antibody fragment to a FATP to occur, such as Fv, Fab, Fab' and F(ab')<sub>2</sub> fragments. Such fragments can be produced by enzymatic cleavage or by recombinant techniques. For instance, papain or pepsin cleavage can generate Fab or F(ab')<sub>2</sub> fragments, respectively. Antibodies can also be produced in a variety of truncated forms using antibody genes in which one or more stop codons have been introduced upstream of the natural stop site. For example, a chimeric gene encoding a F(ab')<sub>2</sub> heavy chain portion can be designed to include DNA sequences encoding the CH<sub>1</sub> domain and hinge region of the heavy chain.

Preparation of immunizing antigen (whole cells comprising FATP on the cell surface or purified FATP), and polyclonal and monoclonal antibody production can be performed using any suitable technique. A variety of methods have been described (See e.g., Kohler *et al.*, *Nature*, 256: 495-497 (1975) and *Eur. J. Immunol.* 6: 511-519 (1976); Milstein *et al.*, *Nature* 266: 550-552 (1977); Koprowski *et al.*, U.S. Patent No. 4,172,124; Harlow, E. and D. Lane, 1988, *Antibodies: A Laboratory Manual*, (Cold Spring Harbor Laboratory: Cold Spring Harbor, NY); Chapter 11 In *Current Protocols In Molecular Biology*, Vol. 2 (containing supplements up through Supplement 42, 1998), Ausubel, F.M. *et al.*, eds., (John Wiley & Sons: New York, NY)). Generally, a hybridoma can be produced by fusing a suitable immortal cell line (e.g., a myeloma cell line such as SP2/0) with antibody producing cells. The antibody producing cells, preferably those obtained from the spleen or lymph nodes, can be obtained from animals immunized with the antigen of interest. Immunization of animals can be by introduction of whole cells comprising fatty acid transport protein on the cell surface. The fused cells (hybridomas) can be isolated using selective culture conditions, and cloned by limiting dilution. Cells which produce antibodies with the desired specificity can be selected by a suitable assay (e.g., ELISA).

Other suitable methods of producing or isolating antibodies (including human antibodies) of the requisite specificity can be used, including, for example, methods

which select recombinant antibody from a library (e.g., Hoogenboom *et al.*, WO 93/06213; Hoogenboom *et al.*, U.S. Patent No. 5,565,332; WO 94/13804, published June 23, 1994; and Dower, W.J. *et al.*, U.S. Patent No. 5,427,908), or which rely upon immunization of transgenic animals (e.g., mice) capable of  
5 producing a full repertoire of human antibodies (see e.g., Jakobovits *et al.*, *Proc. Natl. Acad. Sci. USA*, 90: 2551-2555 (1993); Jakobovits *et al.*, *Nature*, 362:255-258 (1993); Lonberg *et al.*, U.S. Patent No. 5,569,825; Lonberg *et al.*, U.S. Patent No. 5,545,806; Surani *et al.*, U.S. Patent No. 5,545,807; and Kucherlapati, R. *et al.*, European Patent No. EP 0 463 151 B1).

10 Another aspect of the invention is a method for directing an agent to cardiac muscle. The differential expression of FATP6 in cardiac muscle but not in other tissue types allows for the specific targeting of drugs, diagnostic agents, tagging labels, histological stains or other substances specifically to cardiac muscle. A targeting vehicle can be used for the delivery of such a substance. Targeting vehicles  
15 which bind specifically to FATP6 can be linked to a substance to be delivered to the cells of cardiac muscle. The linkage can be, for instance, via one or more covalent bonds, or by high affinity non-covalent bonds. A targeting vehicle can be an antibody, for instance, or other compound (e.g., a fatty acid or fatty acid analog) which binds to FATP6 with high specificity.

20 Targeting vehicles specific to the heart-specific protein FATP6 have *in vivo* (e.g., therapeutic and diagnostic) applications. For example, an antibody which specifically binds to FATP6 can be conjugated to a drug to be targeted to the heart (e.g., a cardiac glycoside to treat congestive heart failure, or  $\beta$ -adrenergic agents, sodium channel blockers or calcium channel blockers to treat arrhythmias). A  
25 substance (e.g., a radioactive substance) which can be detected (e.g., a label) *in vivo* can also be linked to a targeting vehicle which specifically binds to a heart-specific protein such as FATP6, and the conjugate can be used as a labeling agent to identify cardiac muscle cells.

Targeting vehicles specific to FATP6 find further applications *in vitro*. For  
30 example, an FATP6-specific targeting vehicle, such as an antibody (a polyclonal preparation or monoclonal) which specifically binds to FATP6, can be linked to a

substance which can be used as a stain for a tissue sample (e.g., horseradish peroxidase) to provide a method for the identification of cardiac muscle in a sample, as can be used in embryology studies, for example.

In a similar manner, an agent can be directed to the liver of a mammal, as  
5 FATP5 is expressed in liver but not in other tissue types. A targeting vehicle which specifically binds to FATP5 can be conjugated to a drug for delivery of the drug to the liver, such as a drug to treat hepatitis, Wilson's disease, lipid storage diseases and liver cancer. As with targeting vehicles specific to FATP6, targeting vehicles specific to FATP5 can be used in studying tissue samples *in vitro*.

10 The invention also relates to compositions comprising a modulator of FATP function. The term "modulate" as used herein refers to the ability of a molecule to alter the function of another molecule. Thus, modulate could mean, for example, inhibit, antagonize, agonize, upregulate, downregulate, induce, or suppress. A modulator has the capability of altering function of its target. Such alteration can be  
15 accomplished at any stage of the transcription, translation, expression or function of the protein, so that, for example, modulation of a target gene can be accomplished by modulation of the DNA or RNA encoding the protein, and the protein itself.

Antagonists or agonists (inhibitors or enhancers) of the FATPs of the invention, antibodies that bind a FATP, or mimetics of a FATP can be employed in  
20 combination with a non-sterile or sterile carrier or carriers for use with cells, tissues or organisms, such as a pharmaceutical carrier suitable for administration to a mammalian subject. Such compositions comprise, for instance, a media additive or a therapeutically effective amount of an inhibitor or enhancer compound to be identified by an assay of the invention and a pharmaceutically acceptable carrier or  
25 excipient. Such carriers may include, but are not limited to, saline, buffered saline, dextrose, water, ethanol, surfactants, such as glycerol, excipients such as lactose and combinations thereof. The formulation can be chosen by one of ordinary skill in the art to suit the mode of administration. The chosen route of administration will be influenced by the predominant tissue or organ location of the FATP whose function is  
30 to be inhibited or enhanced. For example, for affecting the function of FATP4, a preferred administration can be oral or through a tube inserted into the stomach (e.g.,

direct stomach tube or nasopharyngeal tube), or through other means to accomplish delivery to the small intestine. The invention further relates to diagnostic and pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention.

5           Compounds of the invention which are FATPs, FATP fusion proteins, FATP mimetics, FATP gene-specific antisense poly- or oligonucleotides, inhibitors or enhancers of a FATP may be employed alone or in conjunction with other compounds, such as therapeutic compounds. The pharmaceutical compositions may be administered in any effective, convenient manner, including administration by  
10   topical, oral, anal, vaginal, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal, transdermal or intradermal routes, among others. In therapy or as a prophylactic, the active agent may be administered to an individual as an injectable composition, for example as a sterile aqueous dispersion, preferably isotonic.

          Alternatively, the composition may be formulated for topical application, for  
15   example, in the form of ointments, creams, lotions, eye ointments, eye drops, ear drops, mouthwash, impregnated dressings and sutures and aerosols, and may contain appropriate conventional additives, including, for example, preservatives, solvents to assist drug penetration, and emollients in ointments and creams. Such topical formulations may also contain compatible conventional carriers, for example cream  
20   or ointment bases, and ethanol or oleyl alcohol for lotions.

          In addition, the amount of the compound will vary depending on the size, age, body weight, general health, sex, and diet of the host, and the time of administration, the biological half-life of the compound, and the particular characteristics and symptoms of the disorder to be treated. Adjustment and manipulation of established  
25   dose ranges are well within the ability of those of skill in the art.

          A further aspect of the invention is a method to identify a polymorphism, or the presence of an alternative or variant allele of a gene in the genome of an organism (of interest here, genes encoding FATPs). As used herein, polymorphism refers to the occurrence of two or more genetically determined alternative sequences or alleles  
30   in a population. A polymorphic locus may be as small as a base pair. Polymorphic markers include restriction fragment length polymorphisms, variable number of

tandem repeats (VNTR's), hypervariable regions, minisatellites, dinucleotide repeats, trinucleotide repeats, tetranucleotide repeats, simple sequence repeats, and insertion elements such as Alu. The first identified allelic form, or the most frequently occurring form can be arbitrarily designated as the reference (usually, "wildtype")

5 form, and other allelic forms are designated as alternative (sometimes, "mutant" or "variant"). Diploid organisms may be homozygous or heterozygous for allelic forms.

An "allele" or "allelic sequence" is an alternative form of a gene which may result from at least one mutation in the nucleotide sequence. Alleles may result in altered mRNAs or polypeptides whose structure or function may or may not be

10 altered. Any given gene may have none, one, or many allelic forms (polymorphism). Common mutational changes which give rise to alleles are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

15 Several different types of polymorphisms have been reported. A restriction fragment length polymorphism (RFLP) is a variation in DNA sequence that alters the length of a restriction fragment (Botstein *et al.*, *Am. J. Hum. Genet.* 32:314-331 (1980)). The restriction fragment length polymorphism may create or delete a restriction site, thus changing the length of the restriction fragment. RFLPs have

20 been widely used in human and animal genetic analyses (see WO 90/13668; WO 90/11369; Donis-Keller, *Cell* 51:319-337 (1987); Lander *et al.*, *Genetics* 121:85-99 (1989)). When a heritable trait can be linked to a particular RFLP, the presence of the RFLP in an individual can be used to predict the likelihood that the individual will also exhibit the trait.

25 Other polymorphisms take the form of short tandem repeats (STRs) that include tandem di-, tri- and tetra-nucleotide repeated motifs. These tandem repeats are also referred to as variable number tandem repeat (VNTR) polymorphisms. VNTRs have been used in identity and paternity analysis (US 5,075,217; Armour *et al.*, *FEBS Lett.* 307:113-115 (1992); Horn *et al.*, WO 91/14003; Jeffreys, EP

30 370,719), and in a large number of genetic mapping studies.

Other polymorphisms take the form of single nucleotide variations between individuals of the same species. Such polymorphisms are far more frequent than RFLPs, STRs (short tandem repeats) and VNTRs (variable number tandem repeats). Some single nucleotide polymorphisms occur in protein-coding sequences, in which  
5 case, one of the polymorphic forms may give rise to the expression of a defective or other variant protein and, potentially, a genetic disease. Other single nucleotide polymorphisms occur in noncoding regions. Some of these polymorphisms may also result in defective protein expression (e.g., as a result of defective splicing). Other single nucleotide polymorphisms have no phenotypic effects.

10 Many of the methods described below require amplification of DNA from target samples and purification of the amplified products. This can be accomplished by PCR, for instance. See generally, *PCR Technology, Principles and Applications for DNA Amplification* (ed. H.A. Erlich), Freeman Press, New York, NY, 1992; *PCR Protocols: A Guide to Methods and Applications* (eds. Innis, et al.), Academic Press,  
15 San Diego, CA, 1990; Mattila *et al.*, *Nucleic Acids Res.* 19:4967 (1991); Eckert *et al.*, *PCR Methods and Applications* 1:17 (1991); *PCR* (eds. McPherson *et al.*, IRS Press, Oxford); and US 4,683,202.

Other suitable amplification methods include the ligase chain reaction (LCR) (see Wu and Wallace, *Genomics* 4:560 (1989); Landegren *et al.*, *Science* 241:1077  
20 (1988)), transcription amplification (Kwoh *et al.*, *Proc. Natl. Acad. Sci. USA* 86:1173 (1989), self-sustained sequence replication (Guatelli *et al.*, *Proc. Natl. Acad. Sci. USA* 87:1874 (1990), and nucleic acid based sequence amplification (NASBA). The latter two amplification methods involve isothermal reactions based on isothermal transcription, which produce both single stranded RNA (ssRNA) and  
25 double stranded DNA (dsDNA) as the amplification products in a ratio of about 30 or 100 to 1, respectively.

Another aspect of the invention is a method for detecting a variant allele of a human FATP gene, comprising preparing amplified, purified FATP DNA from a reference human and amplified, purified, FATP DNA from a "test" human to be  
30 compared to the reference as having a variant allele, using the same or comparable amplification procedures, and determining whether the reference DNA and test DNA



differ in DNA sequence in the FATP gene, whether in a coding or a noncoding region, wherein, if the test DNA differs in sequence from the reference DNA, the test DNA comprises a variant allele of a human FATP gene. The following is a discussion of some of the methods by which it can be determined whether the

5 reference FATP DNA and test FATP DNA differ in sequence.

Direct Sequencing. The direct analysis of the sequence of variant alleles of the present invention can be accomplished using either the dideoxy chain termination method or the Maxam and Gilbert method (see Sambrook *et al.*, *Molecular Cloning: A*

10 *Laboratory Manual*, 2nd ed., Cold Spring Harbor Press, New York 1989; Zyskind *et al.*, *Recombinant DNA Laboratory Manual*, Acad. Press, 1988)).

Denaturing Gradient Gel Electrophoresis. Amplification products generated using the polymerase chain reaction can be analyzed by the use of denaturing gradient gel electrophoresis. Different alleles can be identified based on the different sequence-

15 dependent strand dissociation properties and electrophoretic migration of DNA in solution (chapter 7 in Erlich, ed. *PCR Technology, Principles and Applications for DNA Amplification*, W.H. Freeman and Co., New York, 1992).

Single-strand Conformation Polymorphism Analysis. Alleles of target sequences can be differentiated using single-strand conformation polymorphism

20 analysis, which identifies base differences by alteration in electrophoretic migration of single stranded PCR products, as described in Orita *et al.*, *Proc. Natl. Acad. Sci. USA* 86:2766-2770 (1989). Amplified PCR products can be generated as described above, and heated or otherwise denatured, to form single-stranded amplification products. Single-stranded nucleic acids may refold or form secondary structures

25 which are partially dependent on the base sequence. The different electrophoretic mobilities of single-stranded amplification products can be related to base-sequence differences between alleles of target sequences.

Detection of Binding by Protein That Binds to Mismatches. Amplified DNA comprising the FATP gene or portion of the gene of interest from genomic DNA, for

30 example, of a normal individual is prepared, using primers designed on the basis of the DNA sequences provided herein. Amplified DNA is also prepared, in a similar

manner, from genomic DNA of an individual to be tested for bearing a distinguishable allele. The primers used in PCR carry different labels, for example, primer 1 with biotin, and primer 2 with  $^{32}\text{P}$ . Unused primers are separated from the PCR products, and the products are quantitated. The heteroduplexes are used in a mismatch detection assay using immobilized mismatch binding protein (MutS) bound to nitrocellulose. The presence of biotin-labeled DNA wherein mismatched regions are bound to the nitrocellulose via MutS protein, is detected by visualizing the binding of streptavidin to biotin. See WO 95/12689. MutS protein has also been used in the detection of point mutations in a gel-mobility-shift assay (Lishanski, A. *et al.*, *Proc. Natl. Acad. Sci. USA* 91:2674-2678 (1994)).

Other methods, such as those described below, can be used to distinguish a FATP allele from a reference allele, once a particular allele has been characterized as to DNA sequence.

Allele-specific probes. The design and use of allele-specific probes for analyzing polymorphisms is described by e.g., Saiki *et al.*, *Nature* 324:163-166 (1986); Dattagupta, EP 235,726, Saiki, WO 89/11548. Allele-specific probes can be designed so that they hybridize to a segment of a target DNA from one individual but do not hybridize to the corresponding segment from another individual due to the presence of different polymorphic forms in the respective segments from the two individuals. Hybridization conditions should be sufficiently stringent that there is a significant difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the alleles. Some probes are designed to hybridize to a segment of target DNA such that the polymorphic site aligns with a central position (e.g., in a 15-mer at the 7 position; in a 16-mer, at either the 8 or 9 position) of the probe. This design of probe achieves good discrimination in hybridization between different allelic forms.

Allele-specific probes are often used in pairs, one member of a pair showing a perfect match to a reference form of a target sequence and the other member showing a perfect match to a variant form. Several pairs of probes can then be immobilized on the same support for simultaneous analysis of multiple polymorphisms within the same target sequence.

Allele-specific Primers. An allele-specific primer hybridizes to a site on target DNA overlapping a polymorphism, and only primes amplification of an allelic form to which the primer exhibits perfect complementarity. See Gibbs, *Nucleic Acid Res.* 17:2427-2448 (1989). This primer is used in conjunction with a second primer  
5 which hybridizes at a distal site. Amplification proceeds from the two primers, resulting in a detectable product which indicates the particular allelic form is present. A control is usually performed with a second pair of primers, one of which shows a single base mismatch at the polymorphic site and the other of which exhibits perfect complementarity to a distal site. The single-base mismatch prevents amplification  
10 and no detectable product is formed. The method works best when the mismatch is included in the 3'-most position of the oligonucleotide aligned with the polymorphism because this position is most destabilizing to elongation from the primer (see, e.g., WO 93/22456).

Gene Chips. Allelic variants can also be identified by hybridization to nucleic  
15 acids immobilized on solid supports (gene chips), as described, for example, in WO 95/11995 and U.S. Patent No. 5,143,854, both of which are incorporated herein by reference. WO 95/11995 describes subarrays that are optimized for detection of a characterized variant allele. Such a subarray contains probes designed to be complementary to a second reference sequence, which is an allelic variant of the first  
20 reference sequence.

The present method is illustrated by the following examples, which are not intended to be limiting in any way.

## EXAMPLES

### Materials and Methods

25 The following Materials and Methods were used in the work described in Examples 1-5.

Sequence Alignment of FATP Clones. The DNA sequence for mouse FATP1 was obtained from the National Center for Biotechnology Information nonredundant database. cDNAs for mmFATP2, 3, 4, and 5 were obtained by screening mouse

expression libraries (purchased from GIBCO/BRL, Rockville, MD) with probes derived from the cloned expressed sequence tags (ESTs) (Research Genetics, Huntsville, AL). Full-length clones were obtained for mmFATP2 and 5 and partial sequences for mmFATP3 and 4. The sequences described herein have been  
5 deposited in the GenBank database (Accession Nos. FATP2, AF072760; FATP3, AF072759; FATP4, AF072758; FATP5, AF072757).

Neither FATP2 nor FATP5 contains an in-frame stop codon upstream of the putative initiator methionine; initiator methionines were assigned by homology with that in mmFATP1 and by the presence of a signal sequence immediately after it. The  
10 *Mycobacterium tuberculosis*, *Caenorhabditis elegans*, and *Saccharomyces cerevisiae* sequences were present in the dbEST database as part of the sequencing projects for these organisms. Sequences were aligned utilizing a ClustalX algorithm and the resulting alignment exported to SeqVu. Homologous amino acid substitutions are boxed in Figure 1 and were determined using the Dayhoff 250 method with a 50%  
15 homology cutoff.

Cell Transfection and LCFA Uptake. COS cells were cotransfected using the DEAE-dextran method with the mammalian expression vector pCDNA 3.1 (Invitrogen, Carlsbad, CA) expressing the gene for CD2 (pCDNA-CD2) in combination with either a pCDNA 3.1 or pCMVSPORT2 (GIBCO/BRL, Rockville,  
20 MD) expression vector containing one of the murine or nematode *FATP* genes (*pCDNA-mmFATP1*, *pCDNA-FATP2*, *pCMVSPORT-FATP5*, *pCDNA-ceFATPb*). Two days after transfection, cells were assayed for CD2 expression with a phycoerythrin-coupled anti-CD2(PE-CD2) monoclonal antibody (PharMingen, Franklin Lakes, NJ), and fatty acid uptake was assayed with a BODIPY-labeled fatty  
25 acid analogue (Molecular Probes). Briefly, cells were washed twice with PBS (phosphate buffered saline) and stained with PE-CD2 at 4°C for 30 min in PBS containing 10% fetal calf serum. They were then washed three times with PBS/fetal calf serum for 5 min followed by an incubation for 2 min at 37°C in fatty acid uptake solution, which contained 0.1  $\mu$ M BODIPY-FA and 0.1% fatty acid-free BSA  
30 (bovine serum albumin) in PBS (Schaffer, J.E. & Lodish, H.F. (1994) *Cell* 79:427-436). After 2 min, the cells were washed four times with ice-cold PBS/0.1% BSA.

The cells were then removed from the plates with PBS containing 5 mM EDTA and resuspended in PBS containing 10% fetal calf serum and 10 mM EDTA. PE-CD2 and BODIPY-FA fluorescence were measured using a FACScan (Becton Dickinson, Franklin Lakes, NJ). COS cells were gated on forward scatter (FSC) and side scatter (SS). Cells exhibiting more than 300 CD2 fluorescence units (dsim) representing 15% of all cells were deemed CD2 positive and their BODIPY-FA fluorescence was quantitated.

*E. coli*-Based LCFA Uptake Assay. The full-length coding region of mtFATP and a control protein, the mammalian transcription factor TFE3, were subcloned into the inducible, prokaryotic expression vector pET (Novagen, Madison, WI). Expression was induced with 1 mM isopropyl  $\beta$ -D-thiogalactoside (IPTG) for 1 hour, or cells were left uninduced. Cells were washed in PBS/0.1% BSA and resuspended in 1 ml PBS/0.1% BSA containing 0.1  $\mu$ M [ $^3$ H]palmitate (NEN) at 37°C. Uptake was stopped after the indicated incubation time by transferring the cells onto filter paper using a cell harvester (Brandel, Bethesda, MD). Filters were washed extensively with ice-cold PBS/0.1% BSA, and [ $^3$ H]palmitate was quantitated by scintillation counting.

Northern Blots. Northern blot analysis of murine FATP expression was done using poly(A) mRNA blots (Clontech, Palo Alto, CA). Probes of each of the FATPs were derived from the 3' untranslated regions of each gene and were <60% identical in sequence. Probes were labeled by random priming (Boehringer Mannheim, Indianapolis, IN) and hybridized at 65°C. Blots were extensively washed in 0.2% SSC/0.1% SDS at 65°C.

Generation of Phylogenetic Trees. Complete and partial sequences for *FATP* genes from human, rat, mouse, puffer fish, *Drosophila melanogaster*, *C. elegans*, *S. cerevisiae*, and *M. tuberculosis* were aligned using ClustalX. A homologous region of 48 amino acids (residues 472-519 in mmFATP1) from all of the genes was used to determine phylogenetic relationship within ClustalX. Based on these data a phylogenetic tree was generated using Tree View PPC (Figure 5).

Nomenclature. It is proposed that the *FATP* genes be given a species specific prefix (mm, *Mus musculus*; hs, *Homo sapiens*; mt, *M. tuberculosis*; dm, *D.*

*melanogaster*; *ce*, *C. elegans*, *sc*, *S. cerevisiae*) and numbered such that mammalian homologues in different species share the same number but differ in their prefix. Since the two *C. elegans* genes cannot be paired with a specific human or mouse FATP, they have been designated *ceFATPa* and *ceFATPb*.

5    Example 1: Identification of Novel Mammalian FATPs

- The National Center for Biotechnology Information EST database was screened, using the mouse FATP protein sequence (mmFATP1), to identify novel FATPs. This strategy led to the identification of more than 50 murine EST sequences which could be assembled into five distinct contiguous DNA sequences (contigs).
- 10    One contig was identical to the previously cloned FATP, which has been renamed FATP1. Another, which has been renamed FATP2, is the murine homologue of a rat gene previously identified by others as a very long chain acyl-CoA synthase (Uchiyama, A., Aoyama, T., Kamijo, K., Uchida, Y., Kondo, N., Orii, T. & Hashimoto, T. (1996) *J. Biol. Chem.* 271:30360-30365). The other three contigs
- 15    represented novel genes (*FATP3*, *4*, and *5*). Full-length clones for *FATP2* and *FATP5* and nearly complete sequences for *FATP3* and *4* (Figure 1) were obtained by screening cDNA libraries made from mouse day 10.5 embryos and adult liver. Also identified were human homologues for each of the murine genes in the EST database. A sixth human gene was also identified; whether this gene is also present in the
- 20    mouse will require additional studies. Map positions are given in Tables 2 and 3.

          The genetic loci for all of the human genes, with the exception of *FATP5* which was already mapped as an unknown EST, were determined using the radiation hybrid

panels. The map positions given below show the distance (in centiRays) from the

25    closest framework marker. As a guideline, there are approximately 300 kb/cR.

Table 2. Mapping Data for Human Genes

	hsFATP1	Chromosome Chr19 places 13.35 cR from WI-6344 (lod>3.0)
5	hsFATP2	Chromosome Chr15 places 4.92 cR from D15S126 (lod>3.0)
	hsFATP3	Chromosome Chr1 places 13.24 cR from WI-2862 (lod>3.0)
	hsFATP4	Chromosome Chr9 places 7.80 cR from WI-9685 (lod>3.0)
10	hsFATP5	unknown EST previously mapped to near D19S418
	hsFATP6	Chromosome Chr5 places 1.41 cR from WI-4907 (lod>3.0)

The mouse map is an internal backcross panel consisting of 188 mouse backcross DNA's plus 4 controls (B6, Spretus, F1, Water). The backcross was  
 15 constructed by crossing B6 by Spretus animals and then crossing those F1's back to B6. Mapping is accomplished by taking advantage of recombinational events during meiosis, and the use of PCR primers to detect the differences (by size or re-annealing events) at any given locus between the B6 and Spretus allele.

For the purposes of mapping, a novel set of primers (gene of interest)  
 20 is used to amplify from all 188 DNA's and then typed as being a B6 ("B") or a Spretus ("S"). This string of B's and S's is entered into the Map Manager program, which does a best fit calculation by comparing the string of 188 typings from the gene of interest to all loci already extant in the panel, for all 20 chromosomes. The gene of interest is then assigned to a particular area on a particular chromosome according to  
 25 a number of parameters, including the minimalization of double cross-overs, and the highest LOD scores. Indicated in Table 3 are distances to the closest markers on either side of the FATP locus.

Table 3. Mapping Data for Mouse Genes

	mmFATP1	Chromosome 8 places 2.82 cM from D8Mit132 (lod 43.4) and 1.81 cM from D8Mit74 (lod 43.5)
5	mmFATP2	Chromosome 2 places 1.29 cM from D2Mit258 (lod 47.9) and 1.75 cM from D2NDS3 (lod 44.9)
	mmFATP3	Chromosome 3 places 2.54 cM from D3Mit22 (lod 29.5) and 19.62 cM from D3Mit42 (lod 13.6)
10	mmFATP4	Chromosome 2 places 13.78 cM from D2Mit1 (lod 22.9) and 3.85 cM from D2Mit65 (lod 41.9)
	mmFATP5	Chromosome 7 places 7.28 cM proximal of D7Mit21 (lod 28.3)
15		

## Example 2: Assessment of Function

The ability of the newly identified mouse genes to function as fatty acid transporters was assessed using a fluorescence-activated cell sorting-based assay. COS cells were transiently cotransfected with expression vectors encoding the cell surface protein CD2 and either mmFATP1, mmFATP2, or mmFATP5, respectively. Two days after transfection, COS cells were stained with an antibody to CD2 and then incubated with a BODIPY-labeled fatty acid [BODIPY-FA, (Schaffer, J.E. & Lodish, H.F. (1994) *Cell* 79:427-436)]. The cells were then washed extensively, lifted off the dish, and analyzed by fluorescence-activated cell sorting. As judged by the number of CD2-positive cells, the transfection efficiency was approximately 20-30%. Fatty acid uptake was quantitated in the transiently transfected COS cells by measuring the BODIPY-FA fluorescence of the CD2-positive cells. Expression of CD2 had no effect on fatty acid uptake as shown by the finding that COS cells expressing only the transfected CD2 cDNA (CD2-positive) had the same low level of



BODIPY-FA uptake as did untransfected (CD2-negative) control cells (Figure 2A, control). In COS cells cotransfected with CD2 and mmFATP1, mmFATP2, or mmFATP5, uptake of BODIPY-FA by the transfected (CD2-positive) cells was increased between 15- to 90-fold over control (CD2 cDNA only) cells (Figures 2A-  
5 2D).

### Example 3: Expression Patterns of Murine FATPs

Expression patterns of members of the murine *FATP* gene family were characterized by Northern blot analysis; to avoid cross-hybridization, the probes used were from the 3' untranslated region of these genes, which are less than 60% identical  
10 in sequence. The expression pattern of FATP1 agrees with that previously found (Schaffer, J.E. & Lodish, H.F. (1994) *Cell* 79:427-436). Here, expression was seen primarily in heart and kidney. FATP2 is expressed almost exclusively in liver and kidney, which corresponds to the reported tissue distribution of the rat homologue [very long chain acyl-CoA (VLACS)] as assessed by Western blotting (Uchiyama, A.,  
15 Aoyama, T., Kamijo, K., Uchida, Y., Kondo, N., Orii, T. & Hashimoto, T. (1996) *J. Biol. Chem.* 271:30360-30365). FATP3 is present in lung, liver, and testis. FATP5 is expressed only in liver and cannot be detected in other tissues even when the blot is overexposed. The human homologue of FATP5 is also liver specific and is not expressed in a wide array of other tissues tested, including fetal liver.

### 20 Example 4: FATPs Are Evolutionarily Conserved

The EST database was searched, using sequences conserved among the five murine *FATP* genes, for *FATP* genes in other organisms. Two homologues were found in *C. elegans* and one in *M. tuberculosis*. One of the *C. elegans* genes was cloned from a cDNA library and expressed in COS cells, as described for the murine  
25 *FATPs*. Overexpression of the nematode *FATP* resulted in a 15-fold increase of BODIPY-FA uptake compared with control cells (Figure 3). The mycobacterial *FATP* gene was isolated from a phage library and assessed for its ability to facilitate fatty acid uptake. *E. coli* transformed with a prokaryotic, isopropyl  $\beta$ -D-thiogalactoside-inducible expression vector containing the mycobacterial *FATP* gene

demonstrated a significant increase in the rate of [<sup>3</sup>H]palmitate uptake after induction, compared with uninduced bacteria or *E. coli* transformed with a control protein (Figure 4). Novel *FATP* genes were also identified in *F. rubripes* (puffer fish) and *D. melanogaster*.

#### 5 Example 5: Phylogenetic Tree of FATPs

Faergeman *et al.* (Faergeman, N.J., DiRusso, C.C., Elberger, A., Knudsen, J. & Black, P. N. (1997) *J. Biol. Chem.* 272:8531-8538) identified three regions of very strong conservation between the *scFATP* and *mmFATP1* genes. The sequences of the FATPs were compared over a 311-amino acid FATP "signature sequence" which  
10 includes these conserved regions corresponding to amino acids 246-557 in *mmFATP1* (underlined in Figure 1). When compared with the National Center for Biotechnology Information nonredundant database, only one region of the "FATP signature sequence" shows significant homology to other proteins. This small stretch of amino acids (underlined in Fig. 1) is an AMP-binding motif found in a multitude  
15 of other proteins, such as acyl-CoA synthase, several CoA lipases, and gramicidin S synthetase component II (Schaffer, J.E. & Lodish, H.F. (1994) *Cell* 79:427-436). The relevance of this motif to fatty acid transport is unclear. Other highly conserved regions among the FATPs, including long stretches of amino acids >90% identical from mycobacteria to humans, are not found in any other class of proteins. A 48-  
20 amino acid segment of the FATP signature sequence was used to construct a phylogenetic tree (Figure 5). Each of the human and mouse genes form their own branch; *hsFATP6*, which as yet has no murine homologue, is most closely related to *hsFATP3* and *mmFATP3*. As expected, *mVLACS* is closer in sequence to *mmFATP2* than to *hsFATP2*. The *FATP* genes of invertebrates i.e., *C. elegans* and  
25 *D. melanogaster*, are most closely related to each other. Surprisingly, the mycobacterial gene is more closely related to the human and mouse *FATP5* genes than to the FATPs of any of the lower organisms. Whether this reflects coevolution of the mycobacterial and human genes awaits further study.

## Materials and Methods

The following materials and methods were used in the work described in Examples 6-10.

### Isolation of full-length human FATP1 and 4

- 5 Full-length clones encoding human FATP1 and human FATP4 were identified by searching databases for sequences similar to murine FATP1-5 coding regions using the BlastX algorithm (Altschul *et al.*, *J. Mol. Biol.* 215: 403-410, 1990).

A concatamer of nucleotide sequences comprising the coding sequences of mmFATP1 (Genbank Accession U15976), mmFATP2, mmFATP3 (SEQ ID NO:6),  
10 mmFATP4 (SEQ ID NO:8) and mmFATP5 (SEQ ID NO:10) was used to search the Millennium database using the BLASTX algorithm. Sequences with a score >150 were evaluated for whether they represented known FATP coding sequences.

- Human clones with similarity to the 5' end of murine FATP sequences were sequenced completely. Clones encoding full-length human FATP1 were obtained  
15 from a heart cDNA library constructed in the mammalian expression vector pMET7 (Tartaglia *et al.*, *Cell*, 83: 1263-1271, 1995). Clones encoding full-length human FATP4 were obtained from a spleen cDNA library constructed in the mammalian expression vector pMET7.

### Isolation of full-length human FATP6

- 20 Several clones encoding human FATP6 were identified by searching public databases as described above. Five clones were analyzed further by restriction digestion and DNA sequencing. One of these clones (Genbank Accession # AA412064) appeared to be full-length and its entire insert was sequenced.

### DNA Sequence Analysis

- 25 Sequences were aligned with the DNASTar program using the Clustal method. Hydrophobicity plots were generated with DNA Strider using the Kyte Doolittle method.

-77-

### In situ hybridization

Tissues were collected from 8 week old C57/B16 mice. Tissues were fresh frozen, cut on a cryostat at 10  $\mu$ m thickness and mounted on Superfrost Plus slides (VWR). Sections were air dried for 20 minutes and then incubated with ice cold 4% paraformaldehyde (PFA)/phosphate buffered saline (PBS) for 10 minutes. Slides were washed 2 times 5 minutes with PBS, incubated with 0.25% acetic anhydride/1 M triethanolamine for 10 minutes, washed with PBS for 5 minutes and dehydrated with 70%, 80%, 95% and 100% ethanol for 1 minute each. Sections were incubated with chloroform for 5 minutes. Hybridizations were performed with  $^{35}$ S-radiolabeled (5x10<sup>7</sup> cpm/ml) cRNA probes generated from the 3' untranslated regions of mouse FATPs by PCR followed by *in vitro* transcription in the presence of 50% formamide, 10% dextran sulfate, 1x Denhardt's solution, 600 mM NaCl, 10 mM DTT, 0.25% SDS and 10  $\mu$ g/ml tRNA for 18 hours at 55°C. After hybridization, slides were washed with 10 mM Tris-HCl pH 7.6, 500 mM NaCl, 1 mM EDTA (TNE) for 10 minutes, incubated in 40  $\mu$ g/ml RNase A in TNE at 37°C for 30 minutes, washed in TNE for 10 minutes, incubated once in 2x SSC at 60°C for 1 hour, once in 0.2x SSC at 60°C for 1 hour, once in 0.2x SSC at 65°C for 1 hour and dehydrated with 50%, 70%, 80%, 90% and 100% ethanol. Localization of mRNA transcripts was detected by dipping slides in Kodak NBT-2 photoemulsion and exposing for 7 days at 4°C, followed by development with Kodak Dektol developer. Slides were counter stained with haematoxylin and eosin and photographed. Controls for the in situ hybridization experiments include the use of a sense probe which showed no signal above background in all cases.

### Northern Blotting

Human mRNA blots were obtained from Invitrogen or Clontech. PCR fragments from the 3' untranslated regions of human FATPs were used as probes. Blots were probed with  $^{32}$ P-labeled DNA probes using the Rapid-Hyb buffer (Amersham, Buckinghamshire, UK) according to the manufacturer's instructions.

Cell transfection and LCFA uptake. COS cells were cotransfected, using lipofectamine (GIBCO BRL, Rockville, MD) according to the manufacturer's

instructions, with the mammalian expression vector pCDNA3.1 (Invitrogen, Carlsbad, CA) expressing the gene for CD2 in combination with a pMET7 expression vector (Tartaglia *et al.*, *Cell*, 83:1263-1271, 1995) containing hsFATP1 (pMET7-hsFATP1) or hsFATP4 (pMET7-hsFATP4) or pMET7 alone. Two days after  
5 transfection, cells were assayed for CD2 expression with a phycoerythrin-coupled anti-CD2 (PE-CD2) monoclonal antibody (PharMingen, Franklin Lakes, NJ), and fatty acid uptake was assayed with a BODIPY-labeled fatty acid analog (Molecular Probes) as described above.

#### Example 6: Determination of Expression of mmFATPs

10 mmFATP4, and to lesser extent mmFATP2, are expressed at high levels in the brush border layer of the small intestine.

Cell transfection and LCFA uptake. COS cells were cotransfected, using lipofectamine (GIBCO BRL, Rockville, MD) according to the manufacturer's instructions, with the mammalian expression vector pCDNA3.1 (Invitrogen,  
15 Carlsbad, CA) expressing the gene for CD2 in combination with a pMET7 expression vector (Tartaglia *et al.*, *Cell*, 83:1263-1271, 1995) containing hsFATP1 (pMET7-hsFATP1) or hsFATP4 (pMET7-hsFATP4) or pMET7 alone. Two days after transfection, cells were assayed for CD2 expression with a phycoerythrin-coupled anti-CD2 (PE-CD2) monoclonal antibody (PharMingen, Franklin Lakes, NJ), and  
20 fatty acid uptake was assayed with a BODIPY-labeled fatty acid analog (Molecular Probes) as described above.

Absorption of dietary fat requires transport of free fatty acids across the apical membrane of epithelial cells in the small intestine. Previous studies suggested that this transport is protein-mediated; however, the transport protein had not yet been  
25 identified. In situ hybridization was performed on each of the three regions of the small intestine -- duodenum, jejunum and ileum -- as well as the colon, using probes from the 3' untranslated regions of mmFATP1, mmFATP2, mmFATP3, mmFATP4 and mmFATP5, to determine whether any of the mouse FATPs are expressed in the small intestine. It was expected that a protein involved in fatty acid absorption would  
30 be expressed in the epithelial cells of the small intestine, but absent from the colon.

Expression of mmFATPs in the jejunum was identical to that in the ileum in all cases. High levels of mmFATP4 mRNA were present in the epithelial cells of the jejunum and ileum, and lower, but significant, amounts were detected in the epithelial cells of the duodenum. Significantly, FATP4 mRNA was absent from other cell  
5 types of the small intestine and no FATP4 mRNA could be detected in any of the cells of the colon. FATP2 mRNA was present in the epithelial cells of the duodenum at a level similar to that of FATP4, but was present at lower levels in the jejunum and ileum. No signals above background were detected for mmFATP1, mmFATP3 and mmFATP5 in any of the intestinal tissues. mmFATP3 and FATP5 were clearly  
10 detectable by in situ hybridization in adult liver and mmFATP1 could be detected in a variety of tissues on a whole embryo in situ, indicating that the FATP1, 3, and 5 probes were working.

mmFATP4 expression is predominant in the small intestine compared to the other organs of the mouse embryo. In the small intestine, FATP4 expression is  
15 limited to differentiated enterocytes, while no signal is detected in the connective tissue or the undifferentiated epithelial cells in the crypts. Differentiated enterocytes are known to be the cells that mediate the uptake of fatty acids. FATP4 is specifically and strongly expressed in the epithelial cells of adult murine duodenum and ileum but not colon. Other FATPs, such as FATP5, are not expressed in the small intestine.  
20 Thus, FATP4 is the major FATP in the mouse small intestine. Given its high level of expression, it is likely that FATP4, and to a lesser extent FATP2, play an important role in the absorption of fatty acids.

mmFATP2, and mmFATP5 are expressed in hepatocytes

Northern analysis of mmFATP2, mmFATP3, mmFATP4 and mmFATP5  
25 showed expression in the liver. To determine whether these proteins are present in hepatocytes or other cells types present in liver homogenates, in situ hybridizations were performed. mmFATP2, and mmFATP5 mRNA was clearly present in hepatocytes, and was not concentrated in other cell types such as endothelial cells or macrophages. No signal above background was detected for mmFATP1 in any of the  
30 cell types in the liver, consistent with the results of the Northern blotting.

-80-

### Example 7: Isolation and Sequence Analysis of Full-length Human FATP1 and Full-length Human FATP4

To identify human cDNA clones encoding FATP family members, Millennium databases were searched for sequences similar to murine FATP1-5 coding regions. Two clones were analyzed in detail; inspection of the entire DNA sequence of these two clones showed that they encode the human orthologs of mmFATP1 and mm FATP4, respectively. These two clones were designated hsFATP1 and hsFATP4, and their DNA and predicted protein sequences are shown in Figures 44A-44C and 45, and 50A-50C and 51. hsFATP1 is predicted to encode a 646 amino acid, 71 kD protein with multiple membrane-spanning domains (Figure 28A). HsFATP4 is predicted to encode a 643 amino acid, 72 kD protein with multiple membrane spanning domains (See Figure 29A). A comparison of the DNA sequences of mouse and human FATP1 and mouse and human FATP4 (Figures 30A-30B and 31A-31B) shows that the mouse and human orthologs are 85% (FATP1) and 87% (FATP4) identical to each other within the coding sequences given in these figures. At the amino acid level, hsFATP1 and hsFATP4 are ~90% identical to their respective mouse orthologs within the coding region shown in these figures (Figures 32 and 33). The sequence identities between mouse and human FATP1 and FATP4 are considerably higher than the ones observed between different FATP family members within one species (~40%-60%) and are present in the N-terminal part of the protein, a region that is poorly conserved between different FATP family members. This high degree of sequence conservation clearly demonstrates that the newly identified human FATPs are orthologs of mouse FATP1 and FATP4 rather than novel FATP family members.

Table 4 is an identity/similarity matrix comparing the amino acid sequences of FATP1 and 4 from human and mouse. This shows that the gene whose sequence is shown in Figure 43A is indeed human FATP4, since it is 91% identical with the murine FATP4 but only 62% identical with the closest related human FATP, which is FATP1.

Table 4				
Identity/Similarity Matrix				
	hsFATP4	mmFATP4	hsFATP1	mmFATP1
hsFATP4	---	93.2	72.3	72.0
mmFATP4	91.0	---	71.2	71.1
hsFATP1	61.9	61.0	---	92.4
mmFATP1	60.7	59.6	89.5	---

#### Example 8: Isolation and Sequence Analysis of Full-length Human FATP6

A search of EST databases identified a set of overlapping human sequences that were similar to FATPs, but did not have a clear mouse ortholog. One of these EST clones was found to encode a full-length cDNA. The entire insert of this clone was sequenced and designated hsFATP6. The DNA and predicted protein sequences of hsFATP6 are shown in Figures 54A-54C and 55. HsFATP6 is predicted to encode a 619 amino acid, 70 kD protein with multiple membrane-spanning domains (Figure 35A). A comparison of the amino acid sequences of hsFATP6 with other human FATPs shows about 37% identity to either hsFATP1 or hsFATP4 (Figure 36). This degree of sequence identity is similar to what is observed between different mouse FATPs. The phylogenetic analysis described above clearly demonstrates that hsFATP6 is a member of the FATP family, but not an ortholog of any of the mouse FATPs. Comparisons were done with "ALIGN" (E. Myers and W. Miller, "Optimal Alignments in Linear Space," *CABIOS* 4:11-17 (1988) using standard settings.

#### Example 9: Tissue Distribution of Human FATPs

The tissue distribution of human FATPs was assessed by Northern blotting. Human FATP3 was expressed in a large variety of tissues. In contrast, human FATP5 was present at high levels in the liver, but was undetectable in all other tissues examined. Thus, both hsFATP3 and hsFATP5 recapitulate the expression



-82-

pattern of their mouse orthologs (see above). HsFATP6 is a novel FATP with no mouse ortholog as yet. Northern blotting shows that hsFATP6 is expressed at high levels in the heart, but is undetectable in other tissues, including skeletal and smooth muscle. This tissue distribution suggests that human FATP6 performs an important  
5 role in energy metabolism in the heart; blocking FATP6-mediated fatty acid transport may therefore be beneficial for a number of heart diseases, e.g., ischemic heart disease.

To identify the major FATP expressed in the human small intestine, Northern blotting was performed on a blot containing mRNA from human stomach, jejunum,  
10 ileum, colon, rectum and lung. hsFATP5 and hsFATP6 were undetectable in any of these tissues. FATP5 is only expressed in liver and FATP6 only in heart. hsFATP2 was weakly expressed in the colon, and an even weaker signal was detectable in jejunum, ileum and lung lanes. hsFATP3 was expressed well in the lung, but was only weakly expressed in the other tissues tested. Importantly, no difference was seen  
15 in the expression of hsFATP3 between small intestine and stomach or colon, suggesting that the expression observed is not related to fatty acid absorption in the small intestine. hsFATP4 was clearly expressed in both jejunum and ileum; expression was significantly lower in the colon and was absent in the stomach. This expression pattern is consistent with a major role for FATP4 in absorption of fatty  
20 acids in the human gut.

#### Example 10: Expression of hsFATP1 and hsFATP4 Promotes Transport of Fatty Acids

COS cells were cotransfected using lipofectamine with the mammalian expression vector pCDNA-CD2 in combination with one of the FATP-containing  
25 expression vectors (pMET7-hsFATP1 or pMET7-hsFATP4) or an insertless expression vector (pMET7, control) as described in Materials and Methods for Examples 6-10. COS cells were gated on forward scatter and side scatter. Cells exhibiting more than 400 CD2 fluorescence units representing ~30% of all cells were deemed CD2-positive. The percent of CD2-positive cells exhibiting a BODIPY-  
30 fluorescence of >300 is plotted for the three different vectors tested (Figure 37).

#### Example 11: Stable Expression of Human FATP4 in 293 Cells

Stable cell lines were generated as follows. A DNA fragment containing the entire hsFATP4 coding sequence as well as 100 nucleotides of 5' and 50 nucleotides of 3' untranslated region was inserted into the vector pIRES-neo (Clontech, Palo Alto, CA) using standard cloning techniques. The resulting construct or a vector control (pIRES-neo) was transfected into 293 cells using the lipofectamine method (Gibco BRL, Rockville, MD) according to the manufacturer's directions. Cells that had taken up the DNA were selected with 1 mg/ml G418 (Gibco BRL, Rockville, MD). Single colonies were picked 1 to 2 weeks after transfection and grown in medium containing 0.8 mg/ml G418. Colonies were screened for the ability to take up fatty acids by measuring uptake of a fluorescently labeled fatty acid (BODIPY-FA). About 40 colonies transfected with the pIRES-neo containing FATP4 and ~20 colonies transfected with pIRES-neo control were analyzed. All 20 of the vector control clones showed amounts of BODIPY-FA uptake similar to each other and to untransfected 293 cells. In contrast, among the 40 FATP4 transfected clones, 3 had a 5- to 10-fold increased BODIPY-FA uptake compared to any of the vector controls, and a large number (~20) showed an approximately two-fold increase in BODIPY-FA levels. This distribution is consistent with FATP4 conferring increased fatty acid uptake in these cells. One of the cell lines with the highest amount of BODIPY-FA uptake was selected to be used for measuring uptake of tritiated fatty acid.

The uptake of tritiated oleate over time by either FATP4 expressing or control cells was assayed over time. Expression of FATP4 increases the rate of fatty acid uptake by over 3-fold, demonstrating that FATP4 is, like the other FATPs, a functional fatty acid transporter (Figure 38).

#### Example 12: Immuno-staining with FATP4-Specific Antiserum

A polyclonal antiserum against the C-terminus of mmFATP4 was raised using a GST-fusion protein having mmFATP4-specific amino acid sequence 552-643 (AVASP...GEEKL). In western blot experiments, the purified antibody reacted strongly with a synthetic peptide matching the C-terminus of mmFATP4, but not with a corresponding region of mmFATP2, mmFATP3, or mmFATP5. The mmFATP4

specific polyclonal antiserum detects, in western blot experiments with enterocyte lysates from 3 different mice, a ~70 kDa protein, which is in accordance with mmFATP4's predicted molecular weight of 72 kDa. The binding is specific for mmFATP4, since it can be completely abolished by preincubation of the antiserum with the GST-fusion peptide used to raise the antibody.

Immunofluorescence experiments were performed using the anti-mmFATP4 antiserum on fresh frozen sections of murine small intestine. The antibody binding demonstrates strong expression of mmFATP4 in enterocytes, confirming the results of the in situ hybridization experiments. At higher magnifications it is apparent that mmFATP4 is expressed at the apical side of the enterocyte, indicating that the transporter is present in the brush border membrane, which is known to mediate the uptake of fatty acids from the intestinal lumen.

Immuno-electron microscopy studies were performed on fresh frozen murine intestinal cells. The gold particles used, appearing as black specks on the electron micrographs, indicate the subcellular localization of mmFATP4 to be on the microvilli of the enterocyte. It can be seen from electron micrographs that mmFATP4 is localized exclusively in membranes, preferentially the apical plasma membrane, confirming that it is indeed a membrane protein.

#### Methods for Immunofluorescence and Immunogold Electron Microscopy

Unfixed mouse small intestine was washed with Hank's buffered salt solution containing 1 mM EDTA, infused with 2.3 M sucrose solution, and embedded in O.C.T., 4583 compound. The material was thick sectioned (15  $\mu$ M - 40  $\mu$ M). The sections were washed in PBS containing 1% BSA and 0.075% glycine to block non-specific binding. Primary and secondary antibodies were diluted in PBS with 10% FCS and incubated for 1h. The sections were mounted in 90% glycerol/PBS containing 1 mg/ml paraphenylenediamine, and examined with a Bio-Rad MRC 600 confocal, mounted on a Zeiss Axioscop.

For the immunogold labeling, the tissue was fixed with 2% paraformaldehyde in PBS for 10 minutes, after which it was cryoprotected by infiltration with 2.3 M sucrose in 0.1 M phosphate buffer (pH 7.4) containing 20% polyvinylpyrrolidone,

and then mounted on aluminum cryo nails and frozen in liquid nitrogen (Tokuyasu, K.T., *J. Microscop.* 143:139-149, 1986). Ultrathin sections were collected on carbon/formvar-coated nickel grids. The primary antibody (anti-FATP4) was diluted in 10% FCS in PBS and incubated overnight at 4° C, followed by donkey anti-rabbit  
 5 IgG-gold (12 nm) (Jackson Labs) for 1h. The sections were stained in 2% neutral uranyl acetate (20 minutes) and absorption stained with 2% uranyl acetate in 0.2% methylcellulose containing 3.2% polyvinyl alcohol. The sections were examined with a Philips EM 410 electron microscope.

Example 13: Inhibition of Fatty Acid Uptake Specific to FATP4 Demonstrated in  
 10 Isolated Mouse Enterocytes

Phosphorothioate derivatives of the following oligonucleotides were synthesized:

	FATP4-AS2	CCCCCACCAGAGAGGCTCC (SEQ ID NO:103)
	FATP4-AS2MM	CCACCCCCGGAAAGCCTGC (SEQ ID NO:104)
15	FATP4-S2	GGAGCCTCTCTGGTGGGGG (SEQ ID NO:105)

FATP4 AS2 is the antisense oligo; it is designed to be complementary to the sequence extending from nucleotide 10 to nucleotide 28 of the mouse FATP4 coding sequence. FATP4-AS2MM is a control oligo; in the oligo every third nucleotide was changed creating mismatches; the overall nucleotide composition is identical to  
 20 FATP4-AS2 (same number of G, A, T, C). FATP4-S2 is the sense control.

Enterocytes were isolated from the small intestine of mice and incubated for 48h in tissue culture (Figure 40) either without oligonucleotides (squares) or with 100  $\mu$ M FATP4 specific sense (circles) or antisense (diamonds) oligonucleotides. The uptake over time of 25  $\mu$ M oleate was then measured. While the FATP4 sense  
 25 oligonucleotide did not significantly influence the uptake, the antisense oligonucleotide inhibited fatty acid uptake by ~50%.

The effect of either FATP4 sense, antisense or mismatch sequence oligonucleotides on the uptake of fatty acids was measured in enterocytes. Isolated enterocytes were incubated with increasing concentrations of FATP4 antisense

oligonucleotides (solid bars in Figure 41), or a mismatch control oligonucleotide with identical nucleotide composition (stippled bars), or with 100  $\mu$ M of the FATP4 sense-oligonucleotide (lined bar). The medium for this incubation was Dulbecco's modified Eagle's medium with 4.5 g/L glucose, 1 mM sodium pyruvate, 0.01 mg/ml human transferrin and 10% fetal bovine serum. After 48 hours of incubation the uptake of oleate by enterocytes was measured over a 5 minute time interval. Measurements were done in quadruplicate. The uptake assay was done in Hank's buffered salt solution with 10 mM taurocholate. Only the enterocytes given FATP4 antisense oligonucleotide showed a concentration dependent decrease of fatty acid uptake, inhibiting it at a 100  $\mu$ M concentration by ~50%. This effect was FATP4 specific, since only the antisense oligonucleotide which can bind to the FATP4 mRNA and block its translation inhibited uptake, but not a control oligonucleotide differing only in the sequence but not the nucleotide content, ruling out a toxic or otherwise nonspecific inhibitory effect of this oligonucleotide due to its chemical composition.

As a further control experiment, the uptake of oleate was measured along with the uptake of methionine in the same cultured enterocytes. Antisense oligonucleotide, mismatch sequence oligonucleotide, or no oligonucleotide was added to a concentration of 100  $\mu$ M to cultures of enterocytes. After incubation for 48 hours, the uptake of both  $^3$ H-labeled oleate and  $^{35}$ S-labeled methionine was assayed. Results are shown in Figure 42. Fatty acid uptake is at the left side of the paired bars; methionine uptake is on the right side of the paired bars. The fact that amino acid uptake was not influenced by the antisense oligonucleotide treatment further supports the conclusion that the antisense oligonucleotide causes a specific reduction in translation of FATP4-specific mRNA.

#### Example 14: mmFATP2 Is Expressed in Proximal Renal Tubule Epithelium

Northern analysis showed that mmFATP1, mmFATP2, and mmFATP4 are present in the kidney. In situ hybridization (methods as for Example 6) was performed to determine which cell type(s) of the kidney these mRNAs are expressed in. mmFATP1 mRNA was present in virtually all cells throughout the kidney with

no obvious preference for a particular cell type. In contrast, mmFATP2 was expressed only in the renal cortex. Within the cortex, expression of mmFATP2 was restricted to the epithelial cells of the proximal renal tubules. The primary function of proximal renal tubule cells is the reabsorption of filtered salts and nutrients (e.g.,  
5 glucose), a process that requires mitochondrial oxidation and that can utilize fatty acids as energy substrates. Based on the localization of mmFATP2, it is possible that mmFATP2 is important for reabsorption in the kidney by allowing uptake of an energy source (fatty acids) from the blood into renal epithelial cells. Alternatively, if fatty acids need to be reabsorbed in the kidney, similarly to glucose, FATP2 could be  
10 involved in the reabsorption of fatty acids. Determination of the subcellular localization of FATP2 will distinguish between these two possibilities.

Table 5. Mouse FATP mRNA Expression

Mouse Probes	mFATP1	mFATP2	mFATP3	mFATP4	mFATP5
E18.5 embryo expression	everywhere, brain = thymus> heart> brown fat, others	liver (hepatocytes)	-	Brain, small intestine, superior cervical ganglion (SCG), dorsal root ganglion (DRG), other regions have lower expression	Mouse Probes
Duodenum	-	villi (surface epithelium)	-	villi (surface epithelium)	-
Jejunum	-	villi (surface epithelium)	-	villi (surface epithelium)	-
Ileum	-	villi (surface epithelium)	-	villi (surface epithelium)	-
Colon	low expression in the crypt	very low level in the crypt	-	-	-
Kidney	cortex and medulla	proximal tubules	-	-	-

Table 5 (continued). Mouse FATP mRNA Expression

Mouse Probes	mFATP1	mFATP2	mFATP3	mFATP4	mFATP5
Liver	-	hepatocytes	hepatocytes	-	hepatocytes
Pancreas	exocrine secretory units or acinar cells; endocrine pancreas (islet) are negative	exocrine secretory units or acinar cells; endocrine pancreas (islet) are negative	-	-	-
Brain	Neuronal expression throughout the brain including hypothalamus	-	-	Neuronal expression throughout the brain including hypothalamus	-
Heart	myocytes	-	-		
Testis	seminiferous tubules	-	seminiferous tubules		
Lung	bronchiole	-	-		
Adipose	adipocyte	adipocyte	-		

## 10 Example 15: Isolation of full-length human FATP3

Full-length clones encoding human FATP3 were identified by searching databases for sequences similar to the murine FATP1-5 coding regions using the BlastX algorithm (Altschul *et al.*, *J. Mol. Biol.* 215: 403-410, 1990). Human clones with similarity to the 5' end of murine FATP sequences were sequenced completely. A clone encoding full-length human FATP3 was



-90-

obtained from a human bone library constructed in the mammalian expression vector pMET7 (Tartaglia, L.A. *et al.*, *Cell* 83: 1263-1271, 1995). To identify human cDNA clones encoding FATP family members, databases were searched for sequences similar to murine FATP1-5 coding regions. One clone was found to encode the human ortholog of mmFATP3 and was designated hsFATP3. The DNA and predicted protein sequences of hsFATP3 are shown in Figures 94A and 94B. hsFATP3 is predicted to encode a 702 amino acid 75.6 kD protein with multiple membrane-spanning domains. A comparison of the DNA sequences of mouse and human FATP3 shows that the mouse and human orthologs are 81% identical to each other within the coding region. At the amino acid level, hsFATP3 is ~86% identical to mm FATP3 within the coding region. The sequence identities between mouse and human FATP3 are considerably higher than those observed between different FATP family members within one species (~40%) and are present in the N-terminal part of the protein, a region that is poorly conserved between different FATP family members.

#### Example 16: Substrate Specificity of Fatty Acid Transport in hsFATP-Transfected Clones

Using a mammalian expression vector, we generated 40 stable 239 cell lines expressing hsFATP4 and 20 cell lines transfected with a control plasmid. The ability of the different cell lines to take up FA, as assessed by uptake assays using the fluorescently labeled Bodipy-palmitate, correlated well with their FATP4 expression levels determined by Western blotting (FIG. 95). All 20 vector control clones showed amounts of Bodipy-FA uptake similar to each other and to untransfected 239 cells. In contrast, among the 40 FATP4 transfected clones, a large number (~20) showed an approximately 2-fold increase in Bodipy-FA uptake compared to any of the vector controls, and three had a 5- to 10-fold increase in Bodipy-FA uptake.

Several of the cell lines with the highest amount of Bodipy-FA uptake as well as isolated primary enterocytes were used to measure the uptake of radiolabeled FAs. Short-term uptake by 293 cells and enterocytes of all FAs tested was linear (FIG. 97). hsFATP4 expression enhanced the rate of palmitate uptake approximately 3 fold over

293 cells transfected with vector alone (FIG. 97) and also accelerated the uptake of oleate but not of linolate, arachidonate, octanoate, butyrate or cholesterol (Table 6). Isolated primary enterocytes showed a similar preference for palmitate and oleate, and absence of transport of arachidonate, octanoate, and butyrate, but displayed a more robust transport of linolate and cholesterol than the transfected 293 cells.

To further characterize the substrate specificity of FATP4, we measured the uptake by stably transfected 293 cells of 5  $\mu$ M Bodipy-FA in the presence of a 20 fold molar excess (i.e., 100  $\mu$ M) of FAs, FA-derivatives and lipid soluble vitamins and hormones. Both saturated and non-saturated fatty acids containing 10 to 26 C atoms strongly competed for uptake of Bodipy-palmitate (FIG. 96 and Table 7) and thus are presumed to be substrates of FATP4. In contrast, fatty acids with eight or fewer C atoms did not compete and thus are presumed not to be FATP4 substrates. Similarly, esters of long chain FAs and other hydrophobic molecules tested had no effect on uptake of Bodipy-palmitate.

#### 15 LCFA Uptake Assays (Methods)

Bodipy-FA uptake assays using FACS were performed, adapted to a 96-well format. LCFA uptake assays with enterocytes or with stably transfected 293 cells were done as follows. Mixed micelles of radiolabeled FA (NEN) and taurocholate (Sigma) in HBS were generated by brief sonication at 37°C. Equal volumes of cells and micelle solution were mixed, resulting in a final FA concentration of 25  $\mu$ M for antisense assays and 10  $\mu$ M for substrate specificity assays. Final taurocholate concentration was 5 mM. Cells were incubated for the indicated amount of time at 37°C. The assay was stopped by transferring the cells onto filter paper followed by extensive washes with ice-cold HBS containing 0.1% BSA using a cell harvester (Brandell). Incorporated oleate was then determined by  $\beta$ -scintillation counting (Beckman).

Table 6

### Uptake of Different Substrates by FATP4 Expressing Cell Lines and Enterocytes

	Fatty Acid	293 Cells Control*	293 Cells Stably Expressing FATP4	FATP4 specific	Enterocytes*
5	Palmitate	564	1695	1131	3036
	Oleate	662	1122	459	117
	Linolate	640	673	33	116
	Arachidonate	3	5	2	0
	Octanoate	0	0	0	5
10	Butyrate	0	50	50	73
	Cholesterol	319	345	26	531

Uptake of different substrates by enterocytes and by control and stable FATP4-expressing 293 cells. The rates of uptake for the indicated fatty acids was measured over 4 min taking measurements every 30 s. All fatty acids were at a concentration of 10  $\mu$ M in HBS containing 5 mM taurocholate.

\*Uptake measured as *pmol/min 10<sup>6</sup> cells*

Table 7

Competition of Bodipy-FA Uptake by FATP4 Expressing Cells

Fatty Acids	Formula	Competition
Butyric Acid	C <sub>4</sub> H <sub>8</sub> O <sub>2</sub>	-

-93-

5	Caproic Acid	$C_6H_{12}O_2$	-
	Caprylic Acid	$C_8H_{16}O_2$	-
	Capric Acid	$C_{10}H_{20}O_2$	++
	Lauric Acid	$C_{12}H_{24}O_2$	++
	Myristic Acid	$C_{14}H_{28}O_2$	++
10	Palmitic Acid	$C_{16}H_{32}O_2$	++
	Stearic Acid	$C_{18}H_{36}O_2$	+
	Oleic Acid	$C_{18}H_{34}O_2$	++
	Linoleic Acid	$C_{18}H_{32}O_2$	++
	Arachidic Acid	$C_{20}H_{40}O_2$	++
	Lignoceric Acid	$C_{24}H_{48}O_2$	++
	Cerotic Acid	$C_{26}H_{52}O_2$	++

## Fatty Acid Derivatives

15	Fatty Acids	Formula	Competition
	Palmitic Acid Methyl Ester	$C_{17}H_{34}O_2$	-
	Stearic Acid Methyl Ester	$C_{19}H_{38}O_2$	-
	Oleic Acid Ethyl Ester	$C_{20}H_{38}O_2$	-
20	Oleic Acid Oley Ester	$C_{36}H_{68}O_2$	-
	Oleoyl CoA	$C_{39}H_{68}N_7O_{17}P_3S$	-
	Cholesteryl Oleate	$C_{45}H_{78}O_2$	-

Table 7 Continued

Competition of Bodipy-FA Uptake by FATP4 Expressing Cells

Lipid-Soluble Vitamins &amp; Hormones

	Fatty Acids	Formula	Competition
	Retinoic Acid (Pro-Vitamin A)	$C_{20}H_{28}O_2$	$\pm$
	Ergocalciferol (Vitamin D2)	$C_{28}H_{44}O_2$	-
	Tocopherol (Vitamin E)	$C_{29}H_{50}O_2$	-
5	3-Phytylamenadione (Vitamin K1)	$C_{31}H_{46}O_2$	-
	Prostaglandin E2	$C_{20}H_{32}O_5$	-

Competition for Bodipy-FA uptake by FATP4 expressing cells by different hydrophobic compounds. The uptake of 5  $\mu$ M Bodipy-FA, C1-Bodipy-C12 was measured in the presence of a 20-fold molar excess (i.e., 100  $\mu$ M) of the indicated fatty acids or fatty acid derivatives. The maximal 100% inhibition was defined as the amount of Bodipy-FA incorporated in the presence of 200  $\mu$ M lauric acid which was on average 18%  $\pm$  5% that of untreated cells.

- 15    -: 0% - 30% inhibition by the indicated substance  
       $\pm$ : 30% - 50% inhibition  
      +: 50% - 70% inhibition  
      ++: 70% - 100% inhibition

## Example 17: Identification and Characterization of the FATP5 Promoter

## METHODS

## BAC Isolation and Luciferase Constructs

An arrayed BAC library was screened by PCR for FATP5 genomic clones.

- 5 PCR primers designed by a program from the Whitehead Institute's Genome Center specifically amplified a single band of the correct size from mouse genomic DNA. Two putative BACs containing the FATP5 genomic sequence were identified and the presence of FATP5 sequence was confirmed by dot hybridization of the BAC with the mmFATP5 cDNA.
- 10 After isolation of positive BACs, large amounts of bacteria were grown and DNA prepared using a Qiagen maxi-prep kit (Qiagen, Venlo, The Netherlands). The BAC was digested with Sac I and ligated into pZero-2 (Invitrogen, Carlsbad, CA). Inserts containing mmFATP5 genomic sequence were identified by screening colony lifts of the ligation with an  $\alpha$ -<sup>32</sup>P-ATP radiolabeled, random primed (Boehringer-
- 15 Mannheim, Indianapolis, IN) mmFATP5 cDNA as a probe. Positive colonies were picked and restriction analysis with Sac I revealed them to contain an identical, large insert of 8-10 kb. Digestion of the Sac I fragment with BstX I yielded three pieces that were subsequently subcloned into pZero and sequenced using an ABI sequencer (Research Genetics). A 1.3 kb piece containing sequence immediately upstream of
- 20 the FATP5 initiator methionine was subcloned into the Xho I and Bgl II sites of the promoter-less pGL3 luciferase reporter vector (Promega Corp., Madison, WI). 7 kb of additional upstream sequence was subcloned into the Xho I and Sac I sites of the prior construct to yield a final construct containing approximately 8 kb of genomic sequence upstream of the initiator methionine. Deletions of the FATP5 promoter
- 25 were constructed using PCR with the 1.3 promoter construct as the template. Products were amplified with primers containing Hind III (5' primer) and Xho I (3' primer) sites using Elongase (Gibco, Rockville, MD). The resulting fragments were cut with Hind III and Xho I and subcloned into the corresponding sites of the promoter-less pGL3 luciferase reporter vector. The internal 30 base pair deletions,
- 30 GC box mutations, and 10 nucleotide linker scan were all created with the

-96-

Quickchange mutagenesis kit (Stratagene, La Jolla, CA) according to the manufacturer's instructions. At least two different bacterial colonies were picked for each construct. The inserts from both colonies were sequenced to check for unintended point mutations and both constructs were assayed for luciferase activity.

## 5 Cell culture, Transfection, and Luciferase Measurements

HepG2, Hep3B, HT1080, 3T3-L1, BOSC, and HACAT cells were grown in DMEM supplemented with 10% fetal calf serum, 1 x penicillin-streptomycin and glutamine (Gibco, Rockville, MD). Mink lung cells were grown in MEM supplemented with 10% fetal calf serum, 1 x minimal essential amino acids, 1 x  
10 penicillin-streptomycin and glutamine. The evening prior to transfection, cells were plated at 50-60% confluence in 24 well dishes. The following morning, cells were placed in 2 mls of fresh media and 250  $\mu$ L of a  $\text{CaPO}_4$  solution (Invitrogen, Carlsbad, CA) containing 2  $\mu$ g of a luciferase reporter construct and 0.5  $\mu$ g of pCMV- $\beta$ -gal was added to the cells. pCMV- $\beta$ -gal constitutively expresses  $\beta$ -galactosidase and  
15 was used to normalize transfection efficiency (Hua et al., 1998). After 12 hours, the cells were washed twice with DMEM and placed in fresh media. Thirty six hours later, the media over the cells was removed and 250  $\mu$ L of 1 x reporter lysis buffer (Promega Corp., Madison, WI) was added. After vigorous shaking for 15 minutes at room temperature, the supernatants were transferred to Eppendorf tubes and briefly  
20 centrifuged to remove particulates. 20  $\mu$ L from these tubes was used for determination of luciferase activity (Promega Corp., Madison, WI) and 20  $\mu$ L was used for the measurement of  $\beta$ -galactosidase activity (Clontech, Palo Alto, CA). All luciferase values were normalized to  $\beta$ -galactosidase to control for transfection efficiency and expressed as relative luciferase units (RLU). For experiments  
25 comparing different cell lines, promoter activity was computed as a fold induction by dividing the RLU activity of either the -8 or -271 promoter constructs by the RLU activity a promoter-less construct. Each data point was done in triplicate and each experiment was repeated a minimum of three times.

-97-

### Northern Blots, Preparation of Nuclear Extracts, and Gel Shift Assays

Human poly-A northern blots were purchased from a commercial vendor (Clontech, Palo Alto, CA) and probed with a piece of the human FATP5 3' untranslated region specific for FATP5. Nuclear lysates from HepG2 and BOSC  
5 cells were essentially prepared according to the method of Hua et al. and stored at -80°C (Hua et al., 1998). Probes for gel shift assays were end labeled using T4 polynucleotide kinase (Boehringer-Mannheim, Indianapolis, IN) and gel purified. Gel shifts were performed at room temperature in 30 µL reactions comprised of 6 µL 5 X binding buffer (100 mM Tris 8.0, 300 mM KCl, 5 mM EDTA, 8 mM MgCl<sub>2</sub>, and  
10 36% glycerol), 0.5 µL of 100 mM DTT, 1 µL of 10 mg/ml BSA, 2 µL of 2 mg/ml poly dI/dC, and 5 µL nuclear lysate. Ten minutes after the addition of nuclear lysate, 40,000 cpm of <sup>32</sup>P-labeled probe were added. After 20 minutes at room temperature, loading dye was added and the reaction run on a 4% non-denaturing gel.

## RESULTS

### 15 Human FATP5 mRNA is only expressed in adult liver

We had previously reported that mmFATP5 mRNA was only expressed in the liver (Hirsch et al., 1998). To determine if the human isoform of FATP5 was also liver specific, we performed northern analysis using a probe from the 3' transcribed but untranslated region of the human gene. Similar to the mouse homolog, hsFATP5  
20 is liver specific. Interestingly, hsFATP5 was not expressed in fetal liver suggesting that it may be developmentally regulated.

### Identification of a FATP5 promoter

We next set out to determine the cis-acting elements responsible for liver specific expression of FATP5. We identified BACs containing the FATP5 genomic  
25 locus and subcloned a 10 kb Sac I fragment which was subsequently sequenced. The Sac I fragment contains approximately 8 kb of genomic sequence upstream of the FATP5 initiator methionine. Blast searches using the 5' end of the Sac I sequence



-98-

revealed that it contained coding sequence for an unknown gene immediately upstream of FATP5. Since the FATP5 promoter is unlikely to overlap the coding sequence of another gene, we hypothesized that the 10 kb Sac I fragment contained the FATP5 promoter. To test this hypothesis, 8 kb of genomic DNA upstream of the translational initiator of FATP was subcloned into the promoter-less pGL3 luciferase reporter vector. This construct was transiently transfected into the HepG2 liver cell line and luciferase activity was determined. The -8 kb piece of DNA resulted in a 35 fold induction of luciferase activity when compared to a pGL3 vector without the FATP5 genomic sequence (FIG. 100). To determine if this activity reflected tissue specific transcription, the -8 kb luciferase reporter construct was transfected into a variety of additional cell types. While promoter activity was also detected in the Hep3b hepatoma cell line, non-liver cell lines did not express luciferase above the level of the promoter-less vector. Thus, the 8 kb upstream genomic element recapitulated liver specific expression *in vitro*.

15 The FATP5 promoter resides within the 261 base pairs upstream of the initiator methionine and requires a single GC box

To determine the cis-acting elements in the -8 kb of genomic sequence responsible for transcriptional activity, serial 5' deletions of the promoter were constructed and transfected into HepG2 cells. Surprisingly, greater than 90% of the -8 kb was dispensable for promoter activity. A construct containing only 261 base pairs upstream of the initiator methionine resulted in promoter activity equivalent to that of the -8 kb construct (FIG. 101). Identical results were obtained when the deletion series was transfected into Hep3b cells (data not shown). We next determined if promoter activity of a small genetic element was tissue specific.

25 Transfection of a construct containing 271 base pairs upstream of the initiator methionine into a variety of cell lines essentially replicated the results of the -8 kb construct in that expression was observed only in liver derived cell lines (FIG. 102).

Since deletion analysis revealed that bases between -261 and -218 were required for promoter activity, we closely examined this region for binding sites of known transcription factors and found the sequence GGGGCGGGG between

-99-

nucleotides -241 and -232 (FIG. 103A). This sequence binds the Sp1 family of transcription factors and is termed a GC box. To determine if the activity of the -271 construct required the GC box, we mutated the GC box. The first construct deleted nucleotides -241 to -222 which removed the GC box and additional downstream  
5 sequence which, although less optimal, might also bind the Sp1 family of transcription factors (SEQ ID NO.: 107). The second construct had three G to A point mutations in the GC box between nucleotides -241 to -232 (SEQ ID NO.: 108). Such mutations had previously been shown to abolish transcriptional activity of GC boxes (Rodenburg et al., 1997). In contrast to the wild type -271 promoter, both of the  
10 mutated constructs were transcriptionally inactive in HepG2 cells (FIG. 103B). Identical results were also obtained in Hep3B cells (data not shown). This suggests that the GC box between -241 to -232 is essential for transcriptional activity of the FATP5 promoter. We next examined whether the sequences necessary for luciferase activity also bound proteins in nuclear extracts from HepG2 cells. Two different  
15 oligonucleotides were used for gel shift analysis. One oligonucleotide (AF-1) contained nucleotides -250 to -230 (SEQ ID NO.: 111) and the other (AF-2) spanned nucleotides -260 to ~-200 (SEQ ID NO.: 109) (FIG. 104). Both oligonucleotides yielded three significant complexes from HepG2 nuclear extracts. All complexes were specific as 100 fold excess of the same unlabeled oligonucleotide could compete  
20 for binding of the radiolabeled oligonucleotide. Mutant AF-1 oligonucleotides containing three point mutations in the GC box did not bind any proteins in HepG2 nuclear extracts or compete for binding of nuclear proteins to the AF-1 or AF-2 oligonucleotides (data not shown). Oligonucleotides AF-1 and AF-2 also bound recombinant Sp1 (Promega Corp, Madison, WI, data not shown). However, nuclear  
25 extract from BOSC cells, a kidney cell line, and HepG2 cells had identical patterns of complex formation (data not shown).

Identification of novel sequences required for transcriptional activity of the FATP5 promoter

While the GC box between nucleotides 241 and 232 is essential for  
30 transcriptional activity, additional sequences downstream of the GC box might also

-100-

be required for transcription. To determine if such sequences existed, we created 30 base pair internal deletions in the ~-271 construct downstream of the GC box. Constructs that had deletions in sequences between 240 and 180 nucleotides upstream of the FATP5 translational initiator had greatly reduced transcriptional activity in HepG2 cells (FIG. 105). To identify the specific sequences within this region required for FATP5 transcription, a 10 nucleotide linker (CTAACAGGAG) (SEQ ID NO.: 113) was exchanged for wild type sequence within the context of the -271 base pair construct (FIG. 106). Inadvertently, the 210 to 200 construct had a single nucleotide insertion and the 190 to 180 construct had a two nucleotide insertion relative to the wild type sequence. However, several other linker constructs that also had equivalent insertions (230 to 220 or 170 to 160 for example) had high levels of luciferase activity. Thus the decrease in luciferase activity in the 190 to 180 and 210 to 200 constructs is due to changes in the nucleotide sequence and not the result of the nucleotide additions. Transfection of these DNA into HepG2 cells revealed two regions important for transcription. Mutating sequences between nucleotides -210 and ~-200 or between nucleotides -190 and -180 drastically reduced luciferase activity (FIG. 106).

In both humans and mice, FATP5 is only expressed in the liver. To determine the promoter elements mediating liver specific transcription, we isolated a BAC encoding the mouse FATP5 genomic locus and sequenced 10 kb upstream of the transcriptional start. Since this 10 kb of genomic DNA did not contain either a TATA box or GC rich regions found in TATA-less promoters, FATP5 may utilize non-canonical sequences for transcription initiation. Unfortunately, attempts to identify the transcriptional start using primer extension were unsuccessful, perhaps due to secondary structure in the 5' UTR. Since we did not unambiguously determine the transcriptional start site, the nucleotide numbering in all of the promoter constructs refers to the distance from the translational start codon.

#### GC box and Sp1 transcription factors

Since another gene was situated approximately 8 kb upstream of the FATP5 initiator methionine, we hypothesized that promoter elements were likely within this

region of DNA. A luciferase reporter construct containing this sequence was transcriptionally active in two liver cell lines but was inactive in cell lines derived from lung, muscle, kidney, skin, or fibroblasts. Deletion analysis of the -8 kb reporter construct revealed that the FATP5 promoter was contained within the 261 nucleotides upstream of the initiator methionine. Promoter activity in this -261 base pair piece required the presence of a single GC box. Gel shift assays with oligonucleotides containing this GC box revealed the presence of three distinct complexes that required a functional GC box for binding. GC boxes bind the Sp1 family of transcription factors and the multiple complexes could reflect the binding of different members of the Sp1 protein family or different post-translational modifications of Sp1 in HepG2 cells (Rödenburg et al., 1997). Although the Sp1 family of transcription factors is widely expressed, Sp1 has been shown to be important for the transcription of several liver specific genes and is upregulated in liver after birth (Rödenburg et al., 1997). In some cases, Sp1 will facilitate the binding of a tissue specific transcription factor to DNA. For example, Sp1 binding to DNA enhances the binding of C/EBP $\beta$  to an adjacent site in the liver specific CYP2D5 promoter (Lee et al., 1994). Since the C/EBP $\beta$  binding site in the CYP2D5 promoter is suboptimal, C/EBP $\beta$  binding to this site requires the presence of Sp1 or nuclear extract. A similar situation could occur in the FATP5 promoter. Although mutations in the 10 nucleotides downstream of the GC box had no effect on luciferase activity, we did not test mutations immediately upstream of the GC box for effects on promoter activity. It is also possible that Sp1 might bind an unknown liver specific transcription factor and recruit it to the FATP5 promoter. Although, there is no experimental evidence for this, Sp1 has recently been shown to bind to a transcriptional activator so additional interacting proteins are possible (Ryu et al., 1999).

#### Other liver specific transcription factors

Alternatively, since the Sp1 gene family is important for the transcription of many genes which are not liver specific, liver specific promoter elements in the FATP5 promoter might be located elsewhere (Boisclair et al., 1993; Rongnoparut et al., 1991; Sorensen and Wintersberger, 1999). Analysis of the sequence downstream

-102-

of the GC box using TFSearch  
(<http://pdap1.trc.rwcp.or.jp/research/db/TFSEARCH.html>) did not reveal any  
additional transcription factor binding sites of relevance (Heinemeyer et al., 1999;  
Heinemeyer et al., 1998). Further, we were unable to visually identify binding sites  
5 for known liver specific transcription factors in this sequence (De Simone and  
Cortese, 1992; Hanson and Reshef, 1997; Lai, 1992). Thus, we looked  
experimentally for additional promoter elements by mutating the sequence  
downstream of the GC box and identified two additional sites downstream of the GC  
box that were essential for FATP5 transcription. The sequences of these sites do not  
10 conform to any known transcription factor binding sites suggesting the either novel  
proteins bind these elements or that these elements bind known proteins in a novel  
manner. Preliminary gel shift data using oligonucleotides spanning these site  
suggests that these two elements may comprise a binding site for a single complex.  
Further additional data suggests that the complex which binds to these two sites  
15 interacts with the GC box 30 base pairs upstream. Interestingly, we noted a  
palindromic sequence equally split between these two sites (FIG. 107). Since many  
transcription factors bind palindromic DNA elements, it is intriguing to speculate that  
these two sequences contribute to the binding site for a novel transcription factor.  
Current investigations are focused on identifying the proteins binding to these novel  
20 elements and how this element interacts with the GC box.

Several studies have shown that the FATP gene family is regulated by a  
variety of substances including LPS, cytokines, insulin, and diet (Frohnert et al.,  
1999; Hui et al., 1998; Memon et al., 1999). Especially intriguing has been a recent  
report that FATP1 is upregulated by PPAR $\alpha$  ligands in liver cell lines (Martin et al.,  
25 1997; Motojima et al., 1998). Since fatty acids may be endogenous activators of  
PPAR's, transcriptional regulation of FATP1 by PPAR's may represent a physiologic  
feedback loop (Gottlicher et al., 1992; Grimaldi et al., 1999; Schoonjans et al., 1996).  
Given that liver also expresses FATP5, it will be interesting to see whether this genes  
is also regulated by PPAR $\alpha$  and the tools developed here should help address this  
30 question.

Several factors make the FATP5 promoter amenable to further study. First, liver specific transcription of FATP5 can be recapitulated using immortalized cell lines *in vitro*. Second, the minimal required promoter element that confers liver specific transcription is very small. Third, transcriptional activity of this promoter is very robust. Thus, further study of the FATP5 promoter may provide additional insight into the mechanisms of liver specific transcription and regulation of the FATP gene family.

#### Example 18:

##### Materials and Methods

Polyclonal antibodies were raised against proteins containing the N-terminal domain of mouse FATP2 or the C-terminal domain of mouse FATP5 fused to glutathione-S-transferase (GS). Tissues for immunofluorescence were collected from 8 week old mice and a 2 year old chimpanzee. Tissues were fresh frozen, cut on a cryostat and mounted on slides. Immunofluorescence was performed as previously described (Stahl et al., 1999). Pictures were taken on a Zeiss confocal microscope.

To determine FATP2 expression in the gall bladder, mouse gall bladder was incubated with anti-FATP2 antibody as the primary antibody and rhodamine-labeled anti-rabbit IG as the secondary antibody. FATP2 antibody clearly stained the gall bladder epithelium, but did not result in significant staining of other cell types.

(Figure 108)

To further study FATP2 expression, chimpanzee liver was costained with anti-FATP2 antibody(green) and anti CD31 antibody(red). CD31 is expressed on endothelial cells and is used as a marker for blood vessels. FATP2 immunoreactivity was present in large patches which overlap with CD31 positive areas, suggesting that FATP2 protein was present in the space of Diss, the area where hepatocytes exchange nutrients with the blood. This implicates FATP2 in the uptake of fatty acids into hepatocytes. In addition to areas which overlap with CD31 immunoreactivity, FATP2 protein was also present on the cell surface of hepatocytes in a small bead

pattern. Immunoelectronmicroscopy of similar sections showed that FATP2 immunoreactivity was localized in the walls of bile caniculi which are formed by the liver cells. (Figure 109) The presence of FATP2 in bile caniculi in the liver as well as its presence in the gall bladder epithelium suggests a role for FATP2 in either  
5 absorption or secretion of fatty acids into the bile. The levels of free fatty acids in the bile have been associated with the frequency of all stone formation.

To further study FATP5 expression, chimpanzee liver was costained with anti-FATP5 antibody(green) and anti CD31 antibody(red). CD31 is expressed on endothelial cells and is used as a marker for blood vessels. FATP5 immunoreactivity  
10 was present in large patches which overlap with CD31 positive areas, suggesting that FATP5 protein was present in the space of Diss, the area where hepatocytes exchange nutrients with the blood. (Figure 110) This implicates FATP5 in the uptake of fatty acids into hepatocytes.

#### Example 19 Identification and Characterization of Human FATP3 Proteins

##### 15 Isolation of additional humanFATP3 clones

An additional clone encoding human FATP3 was identified by searching for sequences similar to murine or human FATP3 coding regions using the BlastX algorithm in a proprietary database, (Altschul, et al, J. Mol. Bio. 215: 403-410, 1990). One clone, which was identified by random library sequencing, is described as  
20 johni003f04 (SEQ ID NO: 116) extends the open reading frame of the hsFATP3 polypeptide sequence by 30 amino acids at the N-terminus when compared to previously discovered sequences. The DNA sequence of this clone is shown in Figures 111A and 111B, and the predicted protein sequence (SEQ ID NO: 117) is shown in Figure 112. The open reading frame of this clone begins at the initial  
25 nucleotide and includes nucleotide 2240. The first ATG is located at nucleotide number 51, resulting in a predicted protein which includes 730 amino acids. An FATP signature sequence (see Hirsch et al., PNAS, 95:8625-8629, 1998) is clearly

present between amino acids 331 and 640 of hsFATP3. Within this signature sequence hsFATP3 is 48% identical to hsFATP1 at the amino acid level. A consensus AMP-binding motif has been identified (amino acid 333-334). Thus, hsFATP3 is clearly a member of the fatty acid family.

## 5 Functional analysis of FATP3 Clones

SEQ ID NO: 116 is contained in the mammalian expression vector pMET7 (Tartaglia, *et al.*, Cell, 83: 1263-1271, 1995). To determine if the protein encoded by this DNA sequence can mediate fatty acid uptake, SEQ ID NO: 116 was transfected into COS cells. Uptake of a BODIPY-labeled fatty acid was determined as described  
 10 in previous experiments (Hirsch, *et al.*, PNAS, 95: 8625-8629, 1998). Transfection with SEQ ID NO: 116 resulted in a dramatic increase in fatty acid uptake when compared to transfection with vector control. In this experiment, CD31 served as a marker for transfected cells. Only CD31 positive cells were considered for analysis (see Hirsch, *et al.*, PNAS, 95: 8625-8629, 1998 for details). The results (Figure 113)  
 15 demonstrate that SEQ ID NO: 116 encodes a functional fatty acid transport protein.

## Tissue Distribution of human FATP3

Polyclonal antibodies were raised by immunizing rabbits with GST fused to the most C-terminal 89 amino acids of mmFATP3 -  
 (RPPQALNLVQLYSHVSENLPYARPRFLRLQESLATTETFKQQKVRMANEGF  
 20 DPSVLSDPLYVLDQDIGAYLPLTPARYSALLSGDLRJ) (SEQ ID NO: 120).  
 Western blotting experiments with murine tissue lysates using the anti-FATP3 antiserum closely confirmed the unique expression pattern of FATP3 as judged by northern blot experiments. This, together with the fact that the serum reacted only weakly with lysates from cell lines expressing either FATP1, -2, -4 or -5, indicates  
 25 that the antibody recognizes preferentially FATP3, but not other FATP family members.

FATP3 protein was detected in mouse liver, spleen, heart, kidney, testis, white adipose tissue, and most notably in the lung. Further FATP3 expression in the lung was examined by immunofluorescence microscopy. 5 to 10  $\mu$ M thick fresh frozen



unfixed sections of murine and chimpanzee lungs were blocked with 10% FCS/1% donkey serum/1% BSA in HBS and incubated overnight with anti-FATP3 serum in blocking solution. After washing the sections Alexa 488 conjugated donkey anti-rabbit secondary antibodies were used to detect bound anti-FATP3 primary antibodies and nuclei were stained TOTO3. In later experiments, chimpanzee lung was incubated with a mixture of rabbit anti-FATP3 and mouse monoclonal anti-CD31 to visualize FATP3 as well as blood vessels. Sections were imaged on a Zeiss LSM510 confocal microscope. Experiments carried out once with mouse and three times with chimpanzee lung tissue showed that FATP3 is present at high levels in type-II pneumocytes, a cell type responsible for secretion of surfactant, a phospholipid-rich film critical for lung function. The exact function of FAT3 in type II pneumocytes is not yet clear. One hypothesis is that FATP3 is responsible for supplying fatty acid substrates for the synthesis of surfactant.

PCR-based experiments showed that the exocrine as well as endocrine pancreas expresses FATP3. This fact was confirmed by immunofluorescence performed as described above for the lung sections, on chimpanzee pancreas which showed FATP3 localized to the plasma membrane of acinar cells and a punctate expression pattern on the plasma membrane and in the cytosol of alpha and beta cells of the pancreatic islands. The identification of a fatty acid transporter in the insulin producing cells of the pancreas has potentially broad implications for the treatment of type II diabetes and obesity. In both diseases, fatty acid levels in the blood are elevated and, in later stages of the disease, lead to diminished insulin secretion by the pancreas due to the induction of apoptosis in insulin-producing beta cells (Shimabukuro, *et al.*, PNAS, 95: 2498-2502, 1988). Blocking fatty acid uptake into the beta cells could possibly prevent apoptosis and maintain insulin secretion thus preventing the progression from obesity to diabetes.

#### Example 20 Identification of a fatty acid binding domain in FATP4

GST fusion proteins were constructed in pGEX for four regions of hsFATP4 (SEQ ID NO: 52; Figure 51) which were generated by PCR and verified by sequencing. The first three fusion proteins were constructed from regions near the N-

-107-

terminal portion of the protein. SP1 (SEQ ID NO:121) contained amino acid residues 43-239 of the hsFATP4 sequence as shown in Figure 114A. This portion of hsFATP4 contains a lipocalin domain (as shown in Figure 117) as well as a number of residues which in hsFATP4 are upstream of the lipocalin domain. SP2 (SEQ ID NO: 122) contained residues 43-290 of the hsFATP4 sequence as shown in Figure 114B. This portion of the hsFATP4 contains a lipocalin domain and an AMP binding domain as well as a number of residues which are upstream of the lipocalin domain. SP3 (SEQ ID NO: 123) contained amino acid residues 125-290 of the hsFATP4 sequence as shown in Figure 114C). This portion of the hsFATP4 contains a lipocalin domain and an AMP binding domain, but does not contain the upstream residues. The fourth fusion protein was constructed from a region at the C-terminal end of the hsFATP4 polypeptide. SP5 contained amino acid residues 417-643 of hsFATP4 polypeptide as show in Figure 114D (SEQ ID NO: 124).

Proteins were expressed in *E. coli* and purified on glutathione affinity beads using standard techniques. To determine fatty acid binding, beads were mixed with 100  $\mu$ M <sup>14</sup>C-labeled fatty acids in mixed micelles with taurocholate (10mM, Sigma) and incubated for 30 minutes at room temperature. The beads were subsequently washed with PBS containing 10mM taurocholate and radioactivity associated with beads was assessed by scintillation counting. A fusion to the C-terminal domain of hsFATP4 (SP5) did not show any oleate (ARC) binding compared to GST protein alone, while 2 N-terminal fusions (SP1 and 2) bound significant amounts of oleate. (Figure 116).

FATTY ACID	SP1	SP2	SP3	SP5	GST
Oleate	25772 $\pm$ 1326	16172 $\pm$ 1639	4206 $\pm$ 631	2413 $\pm$ 186	1511 $\pm$ 525

Similar results were obtained using maltose-binding protein fusions. MBP fusion constructs were generated by digesting the pGEX-SP constructs with EcoRI/XhoI and ligated into pMAL digested with EcoRI/SaII. MBP fusion proteins were expressed in *E. coli* and were purified under non-denaturing conditions following

-108-

the manufacturer's instructions. To determine fatty acid binding, beads were mixed with 100  $\mu$ M  $^{14}$ C-labeled fatty acids in mixed micelles with taurocholate (10mM, Sigma) and incubated for 30 minutes at room temperature. The beads were subsequently washed with HBS containing 10mM taurocholate. The proteins were subsequently eluted from the resin with maltose and the amount of fatty acid binding to MBP-SP1, -2, -3, and -5 was assessed by determining the radioactivity associated with the elute by  $\beta$ -scintillation counting.

Unlike GST fusion proteins, MBP fusion proteins are not self-dimerizing. Further, long-chain fatty acids (such as oleate and palmitate), but not short-chain fatty acids (such as butyrate), were specifically bound by SP1 (Figure 117). This selective binding is consistent with previous reports of the substrate specificity of FATP4 (Stahl, *et al.*, Mol. Cell, 4, 299-308, 1999). The identification of a fatty acid binding domain in FATP4 will be useful in the development of small molecules that inhibit the binding and transport of fatty acids by FATP4 and may provide useful information on the mechanism of fatty acid transport.

#### Results of Fatty Acid Binding

FATTY ACID	Composition	binding to MBP-SP1	binding to MBP-SP5
Oleate	C18H34O2	3968	2800
Palmitate	C16H32O2	4588	844
Arachidonate	C20H40O2	1942	1147
Butyrate	C4H8O2	142	633

These experiments demonstrate that the FATPs of the present invention contain domains that bind various long chain fatty acids. Thus, polypeptides containing these domains can be prepared and utilized to assess the modulation of binding and transport function by a variety of agents. The polypeptides with the highest binding capacities were shown to be those containing a lipocalin domain (such as those shown in Figure 118) with additional upstream residues, such as those

associated with this domain in the N-terminal portion of hsFATP4. Polypeptides containing domains in addition to the lipocalin domain (for example, those containing an AMP binding domain) were also shown to bind fatty acids at significant levels.

Figure 118 contains an alignment depicting the consensus sequences for the six  
5 human FATP, hsFATP1, hsFATP2, hsFATP3, hsFATP4, hsFATP5 and hsFATP6 polypeptides. A lipocalin domain and an AMP binding domain for each polypeptide are both identified and compared. A search using the lipocalin signature sequence [DENG]-X-[DENQGSTARK]-X(0,2)-[DENQARK]-[LIVFY]-{CP}-G-{C}-W-[FYWLRH-X]-[LIVMTA] conducted on a public database ([www.ebi.ac.uk/interpro/](http://www.ebi.ac.uk/interpro/)),  
10 indicated that the lipocalin domains of hsFATP1 and hsFATP4 are identical to the lipocalin signature sequence. In addition, a search directed to identifying sequences having at least 80% identity to the lipocalin signature sequence identified three additional human FATPs, hsFATP3, hsFATP5 and hsFATP6.

15 The following is the result of comparing individual hsFATP protein sequences with the lipocalin domain identified for hsFATP1 and hsFATP4. The comparison was made using the BLAST Network Service at the National Center for Biotechnology Information. (Capitalized AA agree with the lipocalin signature sequence.)

FATP6: 114 to 125 NEpDFVhVWFGL. 76% similarity (SEQ ID NO: 138)  
20 AATGAGCCGGACTTCGTTACGTGTGGTTCGGCCTC

FATP5: 182 to 194 sQAVpaLcMWLGL. 53% similarity (SEQ ID NO: 139)  
TCCCAGGCCGTTCCAGCCCTGTGTATGTGGCTGGGGCTG

FATP4: 134 to 146 ENRNEFVGLWLGM. Identity (SEQ ID NO: 129)  
GAGAACCGCAATGAGTTCGTGGGCCTATGGCTGGGCATG

-110-

FATP3: 221 to 234 IPAGPEFLwLWFGL. 69% similarity (SEQ ID NO: 140)  
CTCCCCGCTGGCCCAGAGTTTCTGTGGCTCTGGTTCGGGCTG

FATP2: 112 to 124 GNEPAYVwLWLGL. 80% similarity (SEQ ID NO: 127)  
GGTAACGAGCCGGCCTACGTGTGGCTGTGGCTGGGGCTG

5 FATP1: 136 to 148 EGRPEFVGLWLGL. Identity (SEQ ID NO: 126)  
GAGGGCCGGCCGGAGTTCGTGGGGCTGTGGCTGGGCCTG

#### REFERENCES

Abumrad, N., Coburn, C., and Ibrahimi, A. (1999). Membrane proteins implicated in long-chain fatty acid uptake by mammalian cells: CD36, FATP and FABPm. *Biochim Biophys Acta* 1441, 4-13.

Berk, P. D., Bradbury, M., Zhou, S. L., Stump, D., and Han, N. I. (1996). Characterization of membrane transport processes: lessons from the study of BSP, bilirubin, and fatty acid uptake. *Semin Liver Dis* 16, 107-20.

Berk, P. D., and Stump, D. D. (1999). Mechanisms of cellular uptake of long chain free fatty acids. *Mol Cell Biochem* 192, 17-31.

Boisclair, Y. R., Brown, A. L., Casola, S., and Rechler, M. M. (1993). Three clustered Sp1 sites are required for efficient transcription of the TATA-less promoter of the gene for insulin-like growth factor-binding protein-2 from the rat. *J Biol Chem* 268, 24892-901.

- De Simone, V., and Cortese, R. (1992). Transcription factors and liver-specific genes. *Biochim Biophys Acta* 1132, 119-26.
- Fitscher, B. A., Elsing, C., Riedel, H. D., Gorski, J., and Stremmel, W. (1996). Protein-mediated facilitated uptake processes for fatty acids, bilirubin, and other  
5 amphipathic compounds [see comments]. *Proc Soc Exp Biol Med* 212, 15-23.
- Frohnert, B. I., Hui, T. Y., and Bernlohr, D. A. (1999). Identification of a functional peroxisome proliferator-responsive element in the murine fatty acid transport protein gene. *J Biol Chem* 274, 3970-7.
- Glatz, J. F., Luiken, J. J., van Nieuwenhoven, F. A., and Van der Vusse, G. J. (1997).  
10 Molecular mechanism of cellular uptake and intracellular translocation of fatty acids. *Prostaglandins Leukot Essent Fatty Acids* 57, 3-9.
- Gottlicher, M., Widmark, E., Li, Q., and Gustafsson, J. A. (1992). Fatty acids activate a chimera of the clofibrilic acid-activated receptor and the glucocorticoid receptor. *Proc Natl Acad Sci U S A* 89, 4653-7.
- 15 Grimaldi, P. A., Teboul, L., Gaillard, D., Armengod, A. V., and Amri, E. Z. (1999). Long chain fatty acids as modulators of gene transcription in preadipose cells. *Mol Cell Biochem* 192, 63-8.
- Hamilton, J. A. (1998). Fatty acid transport: difficult or easy? *J Lipid Res* 39, 467-81.
- Hanson, R. W., and Reshef, L. (1997). Regulation of phosphoenolpyruvate  
20 carboxykinase (GTP) gene expression. *Annu Rev Biochem* 66, 581-611.

- Heinemeyer, T., Chen, X., Karas, H., Kel, A. E., Kel, O. V., Liebich, I., Meinhardt, T., Reuter, I., Schacherer, F., and Wingender, E. (1999). Expanding the TRANSFAC database towards an expert system of regulatory molecular mechanisms. *Nucleic Acids Res* 27, 318-22.
- 5    Heinemeyer, T., Wingender, E., Reuter, I., Hermjakob, H., Kel, A. E., Kel, O. V., Ignatieva, E. V., Ananko, E. A., Podkolodnaya, O. A., Kolpakov, F. A., Podkolodny, N. L., and Kolchanov, N. A. (1998). Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. *Nucleic Acids Res* 26, 362-7.
- 10    Hirsch, D., Stahl, A., and Lodish, H. F. (1998). A family of fatty acid transporters conserved from mycobacterium to man. *Proc Natl Acad Sci U S A* 95, 8625-9.
- Hua, X., Liu, X., Ansari, D. O., and Lodish, H. F. (1998). Synergistic cooperation of TFE3 and smad proteins in TGF-beta-induced transcription of the plasminogen activator inhibitor-1 gene. *Genes Dev* 12, 3084-95.
- 15    Hui, T. Y., Frohnert, B. I., Smith, A. J., Schaffer, J. E., and Bernlohr, D. A. (1998). Characterization of the murine fatty acid transport protein gene and its insulin response sequence. *J Biol Chem* 273, 27420-9.
- Lai, E. (1992). Regulation of hepatic gene expression and development. *Semin Liver Dis* 12, 246-51.
- 20    Lee, Y. H., Yano, M., Liu, S. Y., Matsunaga, E., Johnson, P. F., and Gonzalez, F. J. (1994). A novel cis-acting element controlling the rat CYP2D5 gene and requiring cooperativity between C/EBP beta and an Sp1 factor. *Mol Cell Biol* 14, 1383-94.

- Martin, G., Schoonjans, K., Lefebvre, A. M., Staels, B., and Auwerx, J. (1997). Coordinate regulation of the expression of the fatty acid transport protein and acyl-CoA synthetase genes by PPARalpha and PPARgamma activators. *J Biol Chem* 272, 28210-7.
- 5 Memon, R. A., Fuller, J., Moser, A. H., Smith, P. J., Grunfeld, C., and Feingold, K. R. (1999). Regulation of putative fatty acid transporters and Acyl-CoA synthetase in liver and adipose tissue in ob/ob mice. *Diabetes* 48, 121-7.
- Motojima, K., Passilly, P., Peters, J. M., Gonzalez, F. J., and Latruffe, N. (1998). Expression of putative fatty acid transporter genes are regulated by peroxisome  
10 proliferator-activated receptor alpha and gamma activators in a tissue- and inducer-specific manner. *J Biol Chem* 273, 16710-4.
- Rodenburg, R. J., Holthuisen, P. E., and Sussenbach, J. S. (1997). A functional Sp1 binding site is essential for the activity of the adult liver-specific human insulin-like growth factor II promoter. *Mol Endocrinol* 11, 237-50.
- 15 Rongnoparut, P., Verdon, C. P., Gehnrich, S. C., and Sul, H. S. (1991). Isolation and characterization of the transcriptionally regulated mouse liver (B-type) phosphofructokinase gene and its promoter. *J Biol Chem* 266, 8086-91.
- Ryu, S., Zhou, S., Ladurner, A. G., and Tjian, R. (1999). The transcriptional cofactor complex CRSP is required for activity of the enhancer-binding protein Sp1. *Nature*  
20 397, 446-50.
- Schaffer, J., and Lodish, H. F. (1995). Molecular mechanisms of long-chain fatty acid



uptake. Trends in Cardiovascular Medicine 5, 218-224.

Schaffer, J. E., and Lodish, H. F. (1994). Expression cloning and characterization of a novel adipocyte long chain fatty acid transport protein [see comments]. Cell 79, 427-36.

- 5 Schoonjans, K., Staels, B., and Auwerx, J. (1996). The peroxisome proliferator activated receptors (PPARS) and their effects on lipid metabolism and adipocyte differentiation. Biochim Biophys Acta 1302, 93-109.

- Sorensen, P., and Wintersberger, E. (1999). Sp1 and NF-Y are necessary and sufficient for growth-dependent regulation of the hamster thymidine kinase promoter [In Process  
10 Citation]. J Biol Chem 274, 30943-9.

Stahl, A., Hirsch, D. J., Gimeno, R. E., Punreddy, S., Ge, P., Watson, N., Patel, S., Kotler, M., Raimondi, A., Tartaglia, L. A., and Lodish, H. F. (1999). Identification of the major intestinal fatty acid transport protein [In Process Citation]. Mol Cell 4, 299-308.

- 15 Stremmel, W. (1989). Mechanism of hepatic fatty acid uptake. Journal of Hepatology 9, 374-382.

All references cited herein are incorporated by reference in their entirety.

- While this invention has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled in the art that  
20 various changes in form and details may be made therein without departing from the spirit and scope of the invention as defined by the appended claims.

-115-

## CLAIMS

What is claimed is:

- 5           1.     An isolated nucleic acid comprising the nucleotide sequence of SEQ ID NO.:116 or its complement.
2.     An isolated nucleic acid comprising the coding sequence of SEQ ID NO.: 116.
3.     An isolated nucleic acid which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO.:117 or its complement.
- 10          4.     An isolated nucleic acid which hybridizes under stringency conditions of 6X SSC at 65° C, followed by at least two washes in 0.2X SSC/0.5% SDS at 65° C, to the nucleic acid comprising the nucleotide sequence of SEQ ID NO.:116.
- 15          5.     An isolated nucleic acid consisting of a nucleotide sequence having at least 95% identity to a nucleotide sequence of Claim 1.
6.     An isolated nucleic acid consisting of a nucleotide sequence having at least 90% identity to a nucleotide sequence of Claim 1.
- 20          7.     An isolated nucleic acid encoding a fusion polypeptide, wherein the isolated nucleic acid comprises a nucleotide sequence of SEQ ID NO.:116.

-116-

8. A vector comprising a nucleic acid of Claim 1.
9. A vector comprising a nucleic acid of Claim 2.
10. A vector comprising a nucleic acid of Claim 3.
11. A vector comprising a nucleic acid of Claim 4.
- 5 12. A vector comprising a nucleic acid of Claim 5.
13. A vector comprising a nucleic acid of Claim 6.
14. A vector comprising a nucleic acid of Claim 7.
15. An isolated host cell transfected with the vector of Claim 8.
16. An isolated host cell transfected with the vector of Claim 9.
- 10 17. An isolated host cell transfected with the vector of Claim 10.
18. An isolated host cell transfected with the vector of Claim 11.
19. An isolated host cell transfected with the vector of Claim 12.

-117-

20. An isolated host cell transfected with the vector of Claim 13.
21. An isolated host cell transfected with the vector of Claim 14.
22. A method of producing a polypeptide comprising the step of culturing the host cell of Claim 15 under conditions in which the nucleic acid is expressed, thereby producing the polypeptide.
- 5
23. A method of producing a polypeptide comprising the step of culturing the host cell of Claim 16 under conditions in which the nucleic acid is expressed, thereby producing the polypeptide.
24. A method of producing a polypeptide comprising the step of culturing the host cell of Claim 17 under conditions in which the nucleic acid is expressed, thereby producing the polypeptide.
- 10
25. A method of producing a polypeptide comprising the step of culturing the host cell of Claim 18 under conditions in which the nucleic acid is expressed, thereby producing the polypeptide.
- 15
26. A method of producing a polypeptide comprising the step of culturing the host cell of Claim 19 under conditions in which the nucleic acid is expressed, thereby producing the polypeptide.
27. A method of producing a polypeptide comprising the step of culturing the host cell of Claim 20 under conditions in which the nucleic acid is

-118-

expressed, thereby producing the polypeptide.

28. A method of producing a polypeptide comprising the step of culturing the host cell of Claim 21 under conditions in which the nucleic acid is expressed, thereby producing the polypeptide.
- 5 29. An isolated nucleic acid comprising at least 30 contiguous nucleotides of the nucleotide sequence of SEQ ID NO.:116.
30. An isolated nucleic acid comprising at least 200 contiguous nucleotides of the nucleotide sequence of SEQ ID NO.:116.
- 10 31. An isolated polypeptide comprising the amino acid sequence of SEQ ID NO.:117.
32. An isolated naturally occurring allelic variant of a polypeptide consisting of the amino acid sequence of Claim 31.
33. An isolated polypeptide consisting of an amino acid sequence having at least 95% identity to the amino acid sequence of Claim 31.
- 15 34. An isolated polypeptide consisting of an amino acid sequence having at least 90% identity to the amino acid sequence of Claim 31.
35. An isolated polypeptide encoded by a nucleic acid that hybridizes to a nucleic acid consisting of the nucleotide sequence of SEQ ID NO.:117

-119-

under stringency conditions of 6X SSC at 65° C, followed by at least two washes in 0.2X SSC/0.5% SDS at 65° C.

36. A fusion protein comprising a polypeptide consisting of the amino acid sequence of SEQ ID NO.:117.
- 5 37. The fusion protein of Claim 36, wherein the fusion protein transports fatty acids across a cell membrane or an artificial cell membrane system.
38. An isolated polypeptide comprising at least 15 contiguous amino acid residues of SEQ ID NO.:117.
- 10 39. An isolated polypeptide comprising at least 50 contiguous amino acid residues of SEQ ID NO.:117.
40. An isolated polypeptide comprising at least 360 contiguous amino acid residues of SEQ ID NO.:117.
- 15 41. An isolated polypeptide comprising an amino acid sequence having at least 15 contiguous amino acid residues of SEQ ID NO.:117, wherein the isolated polypeptide transports fatty acids across a cell membrane or an artificial cell membrane.
- 20 42. An isolated polypeptide encoded by a nucleic acid that hybridizes to a nucleic acid consisting of the nucleotide sequence of SEQ ID NO.:116 under stringency conditions of 6X SSC at 65° C, followed by at least

-120-

two washes in 0.2X SSC/0.5% SDS at 65° C.

- 5                   43.    A method for identifying an agent which binds to a protein comprising an amino acid sequence of SEQ ID NO.:117 comprising the steps of contacting the agent with the isolated protein under conditions appropriate for binding of the agent to the isolated protein, and detecting a resulting agent-protein complex.
44.    An agent identified by the method of Claim 43.
- 10               45.    A method for identifying an agent which is an inhibitor of fatty acid uptake by a protein encoded by a polynucleotide comprising a nucleotide sequence which encodes a protein consisting of the amino acid sequence of SEQ ID NO.:117, comprising the steps of:
- a)     maintaining test cells expressing said polynucleotide in the presence of a fatty acid and an agent to be tested as an inhibitor of fatty acid uptake;
- 15               b)     measuring uptake of the fatty acid in the test cells; and
- c)     comparing uptake of the fatty acid in the test cells with uptake of the fatty acid in suitable control cells;
- 20               wherein lower uptake of the fatty acid in the test cells compared to uptake of the fatty acid in the control cells is indicative that the agent is an inhibitor of fatty acid uptake by said protein.
46.    An inhibitor of fatty acid uptake identified by the method of Claim 45.
47.    The method of Claim 45 further comprising the steps of:

- 5
- a) administering the agent to one or more test animals;
  - b) measuring exogenously supplied fatty acids in one or more samples of tissue or bodily fluid from said test animals;
  - c) measuring exogenously supplied fatty acids in one or more comparable samples of tissue or bodily fluid from suitable control animals;
  - d) comparing the fatty acids of b) with the fatty acids of c);
- whereby, lower fatty acids in step b) than in step c) is indicative that the agent is an inhibitor of said protein.

10      48.      An inhibitor of fatty acid uptake identified by the method of Claim 47.

49.      The method of Claim 45, wherein the nucleotide sequence which encodes a protein consists of a nucleotide sequence with 95% identity to a nucleotide sequence which encodes the polypeptide with SEQ ID NO.: 117.

15      50.      A method for identifying an agent which is an inhibitor of a protein encoded by a polynucleotide comprising a nucleotide sequence which encodes a protein comprising the amino acid sequence in SEQ ID NO.: 117 comprising the steps of:

- 20
- (a) introducing into host cells one or more vectors comprising a polynucleotide expressing said protein;
  - (b) culturing a first aliquot of the host cells with fatty acid substrate of said protein and with an agent being tested as an inhibitor of said protein;
  - (c) culturing a second aliquot of the host cells with fatty acid



-122-

substrate of said protein;

- (d) measuring, in the first and second aliquots, uptake of the fatty acid substrate of the host cells;

wherein less uptake of the fatty acid substrate in the first aliquot compared to the second aliquot is indicative that the agent is an inhibitor of said protein.

5

51. An inhibitor of fatty acid uptake identified by the method of Claim 50.

52. The method of Claim 50 further comprising the steps of:

10

- a) administering the agent to one or more test animals;
- b) measuring exogenously supplied fatty acids in one or more samples of tissue or bodily fluid from suitable control animals;
- c) measuring exogenously supplied fatty acids in one or more comparable samples of tissue or bodily fluid from the test animals; and
- d) comparing the fatty acids of the control animals with the fatty acids of the test animals whereby, lower fatty acids in the control animals than in the test animals is indicative that the agent is an inhibitor of said protein.

15

20

53. A method for identifying an agent which binds to a protein comprising an amino acid sequence of SEQ ID NO.:117 comprising the steps of contacting the agent with the isolated protein under conditions appropriate for binding of the agent to the isolated protein, and detecting a resulting agent-protein complex.

-123-

54. A method for identifying an agent which inhibits interaction between an isolated protein comprising an amino acid sequence of SEQ ID NO.:117, and further comprising a ligand of said protein, comprising:

(a) combining:

- 5 (1) said isolated protein;
- (2) the ligand of said protein; and
- (3) a candidate agent to be assessed for its ability to inhibit interaction between said protein of (1) and the ligand of (2), under conditions appropriate for interaction
- 10 between the said protein of (1) and the ligand of (2);

(b) determining the extent to which said protein of (1) and the ligand of (2) interact; and

(c) comparing the extent determined in (b) with the extent to which interaction of said protein of (1) and the ligand of (2) occurs in the absence of the candidate agent to be assessed and under the same conditions appropriate for interaction of said protein of

15 (1) with the ligand of (2);

wherein if the extent to which interaction of said protein of (1) and the ligand of (2) occurs is less in the presence of the candidate agent than in the absence of the candidate agent, the candidate agent is an agent which inhibits interaction between said protein and the ligand of said protein.

20

55. A method for detecting, in a sample of cells, a nucleic acid molecule consisting of a nucleotide sequence with at least 90% sequence identity to SEQ ID NO.:116, comprising:

25

- a) purifying nucleic acid from the cells;
- b) hybridizing 1) purified nucleic acid from the cells to 2) purified nucleic

-124-

acid comprising SEQ ID NO.:116, under conditions that allow hybridization between 1) and 2) if the sequences of 1) and 2) have at least 90% sequence identity; and

- 5           c)     detecting resulting hybrid nucleic acids in the hybridization; wherein, if hybrid nucleic acids are detected at a significant level compared to a suitable control hybridization, then a nucleic acid molecule comprising at least 90% sequence identity to SEQ ID NO: 116, has been detected.

- 10       56.     A method for identifying (1) nucleic acid molecules in fixed cells which specifically interact with a (2) nucleic acid molecule comprising the nucleotide sequence in SEQ ID NO.:116, said method comprising the steps of:

- 15           a)     adding to the fixed cells the nucleic acid molecule comprising a nucleotide sequence in SEQ ID NO.:116;
- b)     incubating the fixed cells under conditions allowing hybridization of (1) with (2);
- c)     removing the nucleic acid molecule of step a) that has not hybridized; and
- d)     detecting hybrid molecules comprising (1) and (2).

- 20       57.     A method for detecting FATP3 in a sample of cells, comprising the steps of adding an agent that specifically binds to FATP3 to the sample, and detecting the agent specifically bound to the FATP3.

58.     The method of Claim 57 wherein the agent is an antibody which specifically binds to FATP3.

-125-

59. A method for detecting FATP3 in a sample of cell lysate, comprising the steps of adding an agent that specifically binds to FATP3 to the sample, and detecting agent specifically bound to the FATP3 .
- 5 60. The method of Claim 59 wherein the agent is an antibody which specifically binds to FATP3.
61. An isolated antibody which binds to a polypeptide having an amino acid sequence consisting of at least 95% amino acid sequence identity with the amino acid sequence of SEQ ID NO.:117.
- 10 62. An isolated antibody which binds to a fatty acid transport protein having the amino acid sequence of SEQ ID NO.:117.
63. A method for detecting, in a sample of cells, a nucleic acid molecule comprising at least 90% sequence identity to SEQ ID NO.:116 comprising:
- 15 a) purifying nucleic acid from the cells;
- b) hybridizing 1) purified nucleic acid from the cells to 2) purified nucleic acid comprising SEQ ID NO.:116 under conditions that allow hybridization between 1) and 2) if the sequences of 1) and 2) have at least 90% sequence identity; and
- 20 c) detecting resulting hybrid nucleic acids in the hybridization; wherein, if hybrid nucleic acids are detected at a significant level compared to a suitable control hybridization, then a nucleic acid molecule having at least 90% sequence identity to SEQ ID NO.:116 has been detected.

64. A method for detecting, in a sample of purified nucleic acid, a nucleic acid molecule comprising at least 90% sequence identity to SEQ ID NO.: 116 comprising:
- 5 a) hybridizing 1) the sample of purified nucleic acid to 2) purified nucleic acid comprising SEQ ID NO.:116 under conditions that allow hybridization between 1) and 2) if the sequences of 1) and 2) have at least 90% sequence identity; and
- 10 b) detecting resulting hybrid nucleic acids in the hybridization; wherein, if hybrid nucleic acids are detected at a significant level compared to a suitable control hybridization, then a nucleic acid molecule having at least 90% sequence identity to SEQ ID NO.:116 has been detected.
65. A method for detecting FATP3 in a sample of cells, comprising the steps of adding an agent that specifically binds to FATP3 to the
- 15 sample, and detecting agent specifically bound to the FATP3.
66. The method of Claim 65 wherein the agent is an antibody which binds to FATP3.
67. A vector comprising a FATP regulatory sequence and at least one targeting sequence directed to the regulatory region of a nucleic acid
- 20 with a nucleotide sequence selected from the group consisting of:
- a) SEQ ID NO.:46
- b) SEQ ID NO.:48
- c) SEQ ID NO.:116
- d) SEQ ID NO.:52

-127-

e) SEQ ID NO.:54 and

f) SEQ ID NO.:56

68. An isolated host cell transfected with a vector of Claim 67.

5

69. A method of producing a polypeptide comprising culturing the host cell of Claim 68 under conditions in which the nucleic acid is expressed, thereby producing the polypeptide.

70. An isolated nucleic acid comprising a nucleotide sequence encoding a functional portion or fragment of a FATP polypeptide comprising a lipocalin domain.

10

71. The isolated nucleic acid of Claim 70 further comprising a nucleotide sequence encoding upstream amino acid residues.

15

72. An isolated nucleic acid comprising a nucleotide sequence encoding a portion or fragment of a FATP protein containing a lipocalin domain, wherein the nucleotide sequence is selected from the group consisting of portions or fragments of:

a) SEQ ID NO.:46

b) SEQ ID NO.:48

c) SEQ ID NO.:116

d) SEQ ID NO.:52

20

e) SEQ ID NO.:54 and

f) SEQ ID NO.:56.

-128-

73. An isolated nucleic acid of Claim 72 further comprising at least about 90 nucleotides of the sequence upstream of the lipocalin domain.
74. A vector comprising a nucleic acid of Claim 73.
75. An isolated host cell comprising the vector of Claim 74.
- 5 76. A method of producing a polypeptide comprising the step of culturing the host cell of Claim 75 under conditions in which the nucleic acid is expressed, thereby producing the polypeptide.
77. A functional portion or fragment of a FATP polypeptide comprising a lipocalin domain.
- 10 78. The FATP polypeptide of Claim 77 further comprising upstream amino acid residues.
79. An isolated polypeptide comprising an amino acid sequence containing a FATP lipocalin domain, wherein the amino acid sequence is selected from the group consisting of portions or fragments of:
- 15 a) SEQ ID NO.:47;
- b) SEQ ID NO.:49;
- c) SEQ ID NO.:117;
- d) SEQ ID NO.:53;
- e) SEQ ID NO.:55; and
- 20 f) SEQ ID NO.:57.

-129-

- 5                   80.    A functional portion or fragment of a FATP polypeptide comprising an amino acid sequence selected from the group consisting of:
- a)     SEQ ID NO.:126;
  - b)     SEQ ID NO.:127;
  - c)     SEQ ID NO.:128;
  - d)     SEQ ID NO.:129;
  - e)     SEQ ID NO.:130; and
  - f)     SEQ ID NO.:131.
- 10               81.    A fusion protein comprising a polypeptide consisting of a FATP polypeptide containing a lipocalin domain.
82.    The fusion protein of Claim 81 further comprising upstream sequences.
83.    The fusion protein of Claim 82, wherein the upstream sequences comprise at least about 30 amino acid residues of an upstream sequence.
- 15               84.    A fusion protein comprising a polypeptide consisting of a FATP polypeptide containing a lipocalin domain, wherein the polypeptide consists of an amino acid sequence selected from the group consisting of portions or fragments of:
- a)     SEQ ID NO.:47;
  - 20       b)     SEQ ID NO.:49;
  - c)     SEQ ID NO.:117;
  - d)     SEQ ID NO.:53;



-130-

e) SEQ ID NO.:55; and

f) SEQ ID NO.:57.

85. The fusion protein of Claim 84 further comprising upstream sequences.

5 86. A method for identifying an agent which binds to a polypeptide, wherein the polypeptide comprises a FATP lipocalin domain, comprising the steps of contacting the agent with the polypeptide under conditions appropriate for binding of the agent to the polypeptide, and detecting a resulting agent-polypeptide complex.

87. The agent identified by the method of Claim 86.

10 88. A method for identifying an agent which binds to a polypeptide, wherein the polypeptide comprises a FATP lipocalin domain and about 30 amino acid residues of an upstream sequence, comprising the steps of contacting the agent with the polypeptide under conditions appropriate for binding of the agent to the polypeptide, and detecting a  
15 resulting agent-polypeptide complex.

89. The agent identified by the method of Claim 88.

20 90. A method for identifying an agent which binds to a polypeptide, wherein the polypeptide comprises a FATP lipocalin domain and consists of an amino acid sequence selected from the group consisting of portions or fragments of:

a) SEQ ID NO.:47;

-131-

- b) SEQ ID NO.:49;
- c) SEQ ID NO.:117;
- d) SEQ ID NO.:53;
- e) SEQ ID NO.:55; and
- 5 f) SEQ ID NO.:57,

comprising the steps of contacting the agent with the polypeptide under conditions appropriate for binding of the agent to the polypeptide, and detecting a resulting agent-polypeptide complex.

91. An agent identified by the method of Claim 90.

10 92. A method for identifying an agent which binds to a polypeptide, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of:

- a) SEQ ID NO.:126;
- b) SEQ ID NO.:127;
- 15 c) SEQ ID NO.:128;
- d) SEQ ID NO.:129;
- e) SEQ ID NO.:130; and
- f) SEQ ID NO.:131,

20 comprising the steps of contacting the agent with the polypeptide under conditions appropriate for binding of the agent to the polypeptide, and detecting a resulting agent-polypeptide complex.

93. An agent identified by the method of Claim 92.

-132-

94. A method for identifying an agent which binds to a polypeptide comprising a FATP lipocalin domain, wherein the polypeptide is encoded by a nucleotide sequence consisting of portions or fragments of:

- 5           a)     SEQ ID NO.:46;  
            b)     SEQ ID NO.:48;  
            c)     SEQ ID NO.:116;  
            d)     SEQ ID NO.:52;  
            e)     SEQ ID NO.:54; and  
10          f)     SEQ ID NO.:56.

comprising the steps of contacting the agent with the polypeptide under conditions appropriate for binding of the agent to the polypeptide, and detecting a resulting agent-polypeptide complex.

95. An agent identified by the method of Claim 94.

15          96. A method for identifying an agent which binds to a polypeptide comprising a FATP lipocalin domain and upstream sequences, wherein the polypeptide is encoded by a nucleotide sequence consisting of portions or fragments of:

- a.     SEQ ID NO.:46;  
20          b.     SEQ ID NO.:48;  
            c.     SEQ ID NO.:116;  
            d.     SEQ ID NO.:52;  
            e.     SEQ ID NO.:54; and  
            f.     SEQ ID NO.:56.

-133-

comprising the steps of contacting the agent with the polypeptide under conditions appropriate for binding of the agent to the polypeptide, and detecting a resulting agent-polypeptide complex.

97. An agent identified by the method of Claim 96.
- 5 98. An isolated nucleic acid sequence comprising the nucleic acid sequence of SEQ ID NO: 113.
99. The portion of the isolated nucleic acid sequence of Claim 98 which encodes a FATP regulatory protein.
- 10 100. The portion of the isolated nucleic acid sequence of Claim 98 which encodes a FATP5 promoter.
101. A method of identifying an agent which alters the level of expression of the nucleic acid encoding an FATP protein comprising:  
determining a base level of expression of the nucleic acid encoding the FATP protein;
- 15 (b) contacting an agent with an isolated nucleic acid containing the coding region of the FATP protein under functional control of its promoter under conditions suitable for binding of the agent to the promoter;
- (c) maintaining agent-promoter binding during expression of the FATP protein; and
- 20 (d) comparing the level of expression of the agent bound promoter to that of the baseline level of expression,

-134-

whereby, if the level of expression of the agent bound promoter is significantly different from that of the baseline level of expression, then an agent which alters the level of expression of the nucleic acid encoding the FATP protein has been identified.

- 5           102.   The method of Claim 101, wherein the FATP protein is FATP2.
103.   The method of Claim 102, wherein the FATP2 is encoded by a nucleic acid comprising the nucleotide sequence of SEQ ID NO: 48.
104.   The method of Claim 102, wherein the FATP2 comprises the amino acid sequence of SEQ ID NO: 49.
- 10          105.   The method of Claim 102, wherein expression is inhibited.
106.   The method of Claim 102, wherein expression is promoted.
107.   The method of Claim 101, wherein the FATP protein is FATP5.
108.   The method of Claim 107, wherein the FATP5 is encoded by a nucleic acid comprising the nucleotide sequence of SEQ ID NO: 54.
- 15          109.   The method of Claim 107, wherein the FATP5 comprises the amino acid sequence of SEQ ID NO: 55.
110.   The method of Claim 107, wherein expression is inhibited.

-135-

111. The method of Claim 107, wherein expression is promoted.
112. A method for directing an agent to liver cells in a mammal, comprising administering to the mammal a complex which comprises the agent and a moiety which binds to FATP2.
- 5 113. The method of Claim 112, wherein the agent alters fatty acid uptake in liver cells.
114. The method of Claim 112, wherein the agent alters the level of fatty acids in bile.
- 10 115. A method for directing an agent to the gall bladder in a mammal, comprising administering to the mammal a complex which comprises the agent and a moiety which binds to FATP2.
116. The method of Claim 115, wherein the agent alters the level of fatty acids in bile.
- 15 117. A method for directing an agent to the liver in a mammal, comprising administering to the mammal a complex which comprises the substance and a moiety which binds to FATP5.
118. The method of Claim 117, wherein the agent alters the uptake of fatty acids in liver cells.

119. The use of an isolated nucleic acid comprising the nucleotide sequence of SEQ ID NO.:116 or its complement in the manufacture of a medicament.
- 5 120. The use of an isolated polypeptide comprising the amino acid sequence of SEQ ID NO.:117 in the manufacture of a medicament.
121. The use of an agent which is an inhibitor of fatty acid uptake of a protein with the amino acid sequence of SEQ ID NO.:117 in the manufacture of a medicament.
- 10 122. The use of an isolated nucleic acid comprising a nucleotide sequence encoding a portion or fragment of a FATP protein containing a lipocalin domain in the manufacture of a medicament, wherein the nucleotide sequence is selected from the group consisting of portions or fragments of:
- 15 a) SEQ ID NO.:46  
b) SEQ ID NO.:48  
c) SEQ ID NO.:116  
d) SEQ ID NO.:52  
e) SEQ ID NO.:54 and  
f) SEQ ID NO.:56.
- 20 123. The use of an isolated polypeptide comprising an amino acid sequence containing a FATP lipocalin domain in the manufacture of a medicament, wherein the amino acid sequence is selected from the

-137-

group consisting of portions or fragments of:

- 5
- a) SEQ ID NO.:47;
  - b) SEQ ID NO.:49;
  - c) SEQ ID NO.:117;
  - d) SEQ ID NO.:53;
  - e) SEQ ID NO.:55; and
  - f) SEQ ID NO.:57.

10 124. The use of an isolated polypeptide in the manufacture of a medicament, the polypeptide comprising an amino acid sequence selected from the group consisting of:

- 15
- 1. SEQ ID NO.:126;
  - 2. SEQ ID NO.:127;
  - 3. SEQ ID NO.:128;
  - 4. SEQ ID NO.:129;
  - 5. SEQ ID NO.:130; and
  - 6. SEQ ID NO.:131.

20 125. The use of an isolated polypeptide in the manufacture of a medicament for treating obesity, the polypeptide comprising an amino acid sequence selected from the group consisting of:

- 7. SEQ ID NO.:126;
- 8. SEQ ID NO.:127;
- 9. SEQ ID NO.:128;
- 10. SEQ ID NO.:129;



-138-

11. SEQ ID NO.:130; and
12. SEQ ID NO.:131.



Fig. 2A

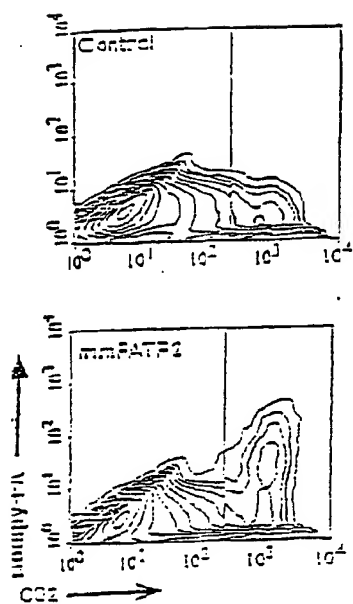


Fig. 2B

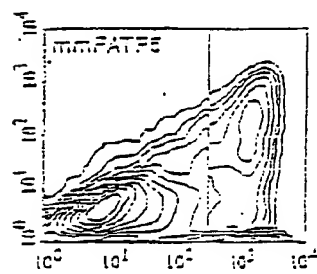
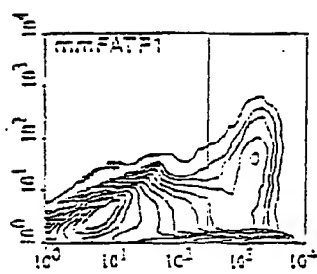


Fig. 2C

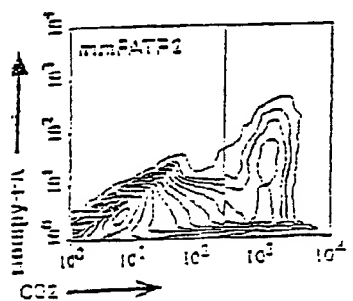


Fig. 2D

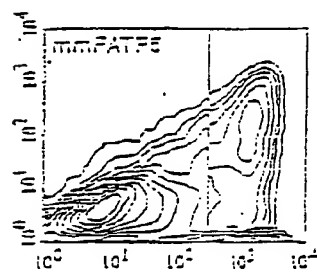


Fig. 3

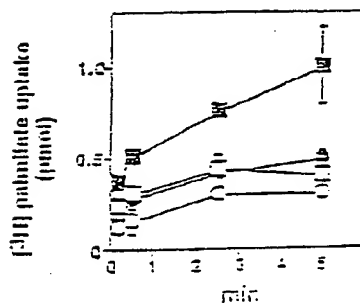
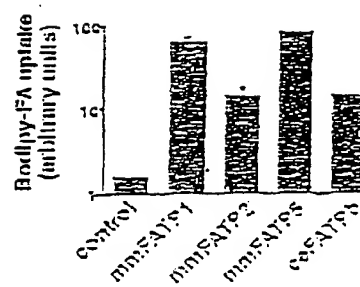


Fig. 4

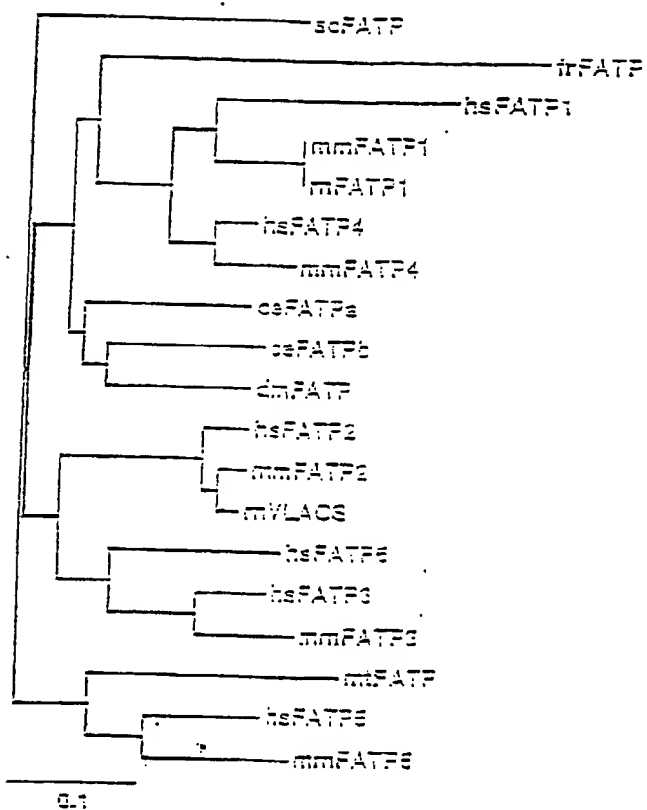


Figure 5

FIG. 6

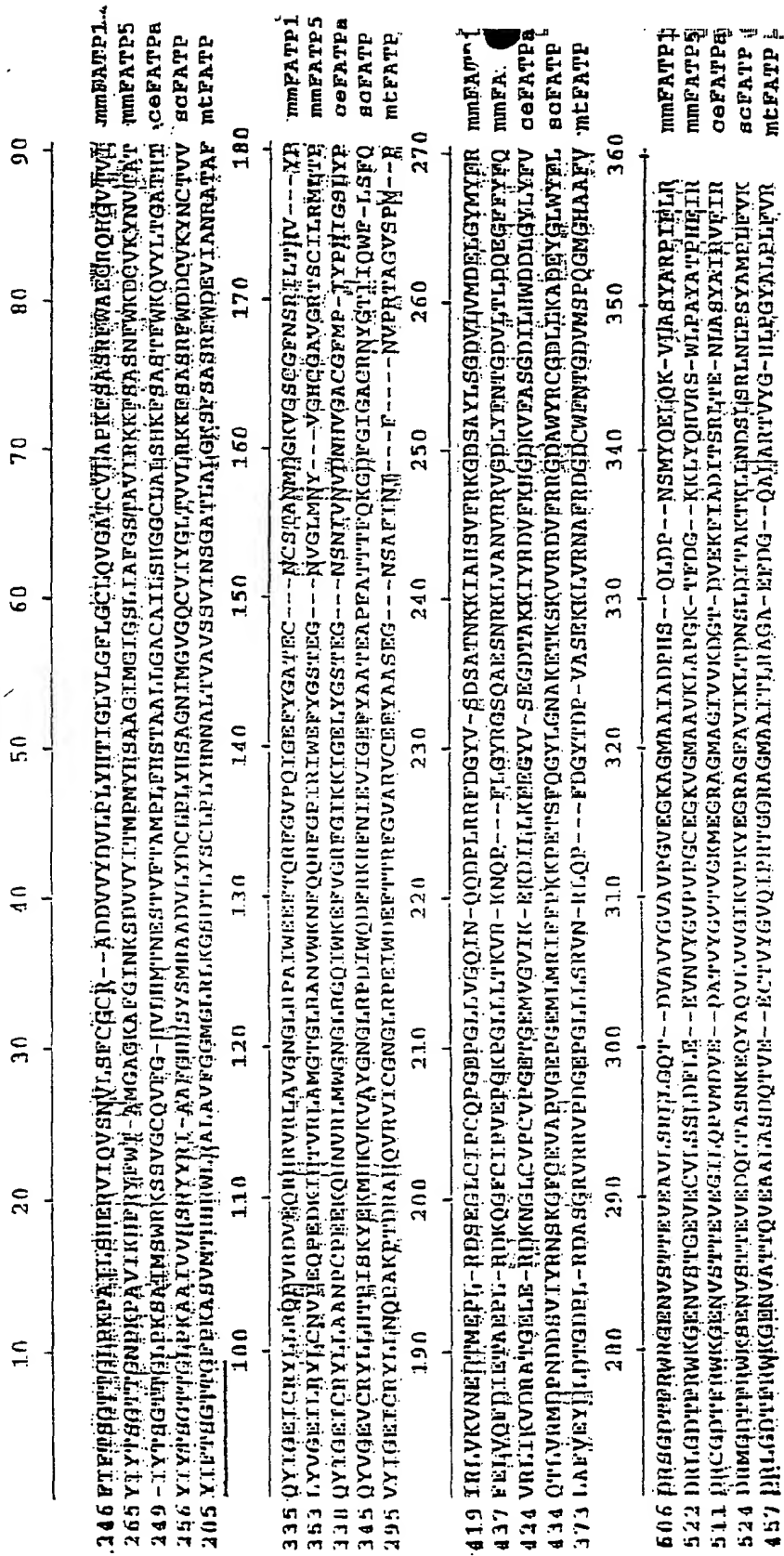


Figure 7



Figure 8B



AADPSSGSSGSIANFLAVIAPFOPTIHELTHAGFSVAFZPSSCIA 50  
 RRIAPFNGTCGPGSGSGSIEGRIAPFAGTAAGNTAFIAPGIV 100  
 ANHEDGFWLWFCAXVETAFVETALFPGIIFPSCPSIWL 150  
 AHSPLSTHETDIPFMEHTNPTECHINAGISNLLSPADQVEVP 200  
 GNSAPQNMDDIOWTISCITEETAPFISTAVLCCQEVHCEHDE 250  
 DVTIALHVMSSGSIIGVCGIGCATVVLKAFSPSQFWLICGHWL 300  
 VEQVIGCGVWNOFSSPSTHAWRIAYGSEFDTWNEFLATFCHQ 350  
 ILHVMGQWAFSTHDSQCMVBSAWLKHIFFSIAPNVMGSH 400  
 ENQCHVMSSEFEINAPVQCSPLGKAPFAHFLAVWWSG 450  
 INFMGGLMDDGELHHPGCHDQKENVATEVAVLCHDFL 500  
 QENAVGAVFEEGAGAAIAPFQAINLWLNSEVSELEFVPH 550  
 FLIQESATETKCHVMAHGGESVLSPLVLRDTEMLPAP 600  
 APYAMSGSH 615

Figure 9

CCG 40  
 CCG 80  
 CCG 120  
 TCG 160  
 CCG 200  
 CCG 240  
 ATCG 280  
 CCG 320  
 ATCG 360  
 TCG 400  
 CCG 440  
 CCG 480  
 CCG 520  
 CCG 560  
 ATCG 600  
 TCG 640  
 CCG 680  
 CCG 720

Figure 10A

[illegible]

Figure 10B

Figure 11

Figure 12A

Figure 12B

MGAIAWFIQPTUWILGAIICWISSAMWLSIVG 40  
 AAIITLILIQPPFLWIFKNAFIFKMFVEKFRRL 80  
 NAGPFRFVDALEQPIAMFIFVAINCHGSSSTINSQ 120  
 DAFSQAPWILZAFKDAVIGNISTPAATINLPSATISAL 160  
 SVTIGAAIGGVWVWIFESPQAFIESVREKASVILVD 200  
 EIIQENLAEVLEKLIADNIEFVIGSSIFVFAIGSL 240  
 DAFSDFEASPAITWESPAIFIFISGTFEIGKILS 280  
 FEFVIGVWMLSPGCAIDVWVWIFVWVWIFVWVW 320  
 GIVGTCVLAFKESASFWAFCHVWVWVWVWVW 360  
 LONVDFPFAIPVFIANGREPMWVWVWVWVWVW 400  
 WEFVGSINGMGMVWVWVWVWVWVWVWVWVW 440  
 FVDFPFAIPVFIANGREPMWVWVWVWVWVWVW 480  
 FGSQFSTFVWVWVWVWVWVWVWVWVWVWVW 520  
 FVDFPFAIPVFIANGREPMWVWVWVWVWVWVW 560  
 GGMVWVWVWVWVWVWVWVWVWVWVWVWVW 600  
 FVDFPFAIPVFIANGREPMWVWVWVWVWVWVW 640  
 AGTFPSMELVWVWVWVWVWVWVWVWVWVW 680

Figure 13

AAGCTTCACTTCTGCTGCTGCTGCTGCTGCTGCT 40  
 AAGCTTCACTTCTGCTGCTGCTGCTGCTGCTGCT 80  
 TTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 160  
 AAGCTTCACTTCTGCTGCTGCTGCTGCTGCTGCT 200  
 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240  
 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 280  
 TTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 320  
 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360  
 TTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 400  
 AAGCTTCACTTCTGCTGCTGCTGCTGCTGCTGCT 440  
 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480  
 TTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 520  
 TTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 560  
 AAGCTTCACTTCTGCTGCTGCTGCTGCTGCTGCT 600  
 AAGCTTCACTTCTGCTGCTGCTGCTGCTGCTGCT 640  
 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 680  
 AAGCTTCACTTCTGCTGCTGCTGCTGCTGCTGCT 720  
 TTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 760  
 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 800

Figure 14A



AAACCTGACCTGAAACCCGCAATGCGGCAAACTA 200  
CGCCGACGACGTCGCGCGGCGGCGGCGGCGGCGG 240  
CGGACGACGTCGCGGCGGCGGCGGCGGCGGCGGCG 280  
CGCCGACGACGTCGCGCGGCGGCGGCGGCGGCGGCG 320  
TTTCGACGACGTCGCGGCGGCGGCGGCGGCGGCGGCG 360  
TTTCGACGACGTCGCGGCGGCGGCGGCGGCGGCGGCG 400  
AAACCTGACCTGAAACCCGCAATGCGGCAAACTA 440  
TTTCGACGACGTCGCGCGGCGGCGGCGGCGGCGGCG 480  
CGCCGACGACGTCGCGCGGCGGCGGCGGCGGCGGCG 520  
TTTCGACGACGTCGCGGCGGCGGCGGCGGCGGCGGCG 560  
AAACCTGACCTGAAACCCGCAATGCGGCAAACTA 600  
AAACCTGACCTGAAACCCGCAATGCGGCAAACTA 640  
TTTCGACGACGTCGCGGCGGCGGCGGCGGCGGCGGCG 680  
CGCCGACGACGTCGCGCGGCGGCGGCGGCGGCGGCG 720  
TTTCGACGACGTCGCGGCGGCGGCGGCGGCGGCGGCG 753

Figure 16B

CGGACGACGTCGCGGCGGCGGCGGCGGCGGCGGCG 40  
CGCCGACGACGTCGCGCGGCGGCGGCGGCGGCGGCG 80  
VWVCGACGACGTCGCGGCGGCGGCGGCGGCGGCGGCG 120  
CGGACGACGTCGCGGCGGCGGCGGCGGCGGCGGCG 160  
CGGACGACGTCGCGGCGGCGGCGGCGGCGGCGGCG 191

Figure 17

TTTCGACGACGTCGCGGCGGCGGCGGCGGCGGCGGCG 40  
CGCCGACGACGTCGCGCGGCGGCGGCGGCGGCGGCG 80  
AAACCTGACCTGAAACCCGCAATGCGGCAAACTA 120  
CGGACGACGTCGCGGCGGCGGCGGCGGCGGCGGCG 160  
CGCCGACGACGTCGCGCGGCGGCGGCGGCGGCGGCG 200  
CGGACGACGTCGCGGCGGCGGCGGCGGCGGCGGCG 240  
TTTCGACGACGTCGCGGCGGCGGCGGCGGCGGCGGCG 280  
TTTCGACGACGTCGCGGCGGCGGCGGCGGCGGCGGCG 320  
CGCCGACGACGTCGCGCGGCGGCGGCGGCGGCGGCG 360  
CGGACGACGTCGCGGCGGCGGCGGCGGCGGCGGCG 400  
CGCCGACGACGTCGCGCGGCGGCGGCGGCGGCGGCG 440  
AAACCTGACCTGAAACCCGCAATGCGGCAAACTA 480  
CGGACGACGTCGCGGCGGCGGCGGCGGCGGCGGCG 520

Figure 18A

Figure 18B

Figure 19

Figure 20A





AATGCGTTCCTTCTTTTAAACAAATGCTCTTCTAT 760  
 TTGCAAAAAGTTTATCACTAATTCCTAATTCTCTAC 800  
 CTTTCTGTTCTCTACGCTTTTAACTAATTCCTAATAAT 840  
 CCACTCTACCAACATTCAACTATTCTAGCTTCTCTTC 880  
 GTCTACCTCTCTTTTATTCCTTAAATTTCTCAACAC 920  
 TTTCTTTATCTCTAATTCCTAATACTCTCTCTCTCT 960  
 CACTCTCTAATTTCTCTCTAATCTCTCTCTCTCTCT 1000  
 AAACTTTCTCTTTCTCTCTCTCTCTCTCTCTCTCT 1040  
 CTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1080  
 AATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1120  
 TTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1160  
 TTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1200  
 CTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1240  
 TTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1280  
 CTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1320  
 AATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1360  
 A 1361

Figure 22B

AATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 50  
 TTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 100  
 CTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 150  
 CTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 200  
 CTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 250  
 CTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 300  
 CTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 335

Figure 23

TTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 40  
 TTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 80  
 CTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120  
 CTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 160  
 CTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 200  
 AATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 240  
 AATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 280  
 AATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 320  
 CTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 360

Figure 24A

TTCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	400
TEGPGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	440
CGTGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	480
TGCTAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	520
TGCTAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	560
CGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	600
CGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	640
CGTGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	680
CGTGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	720
CGTGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	760
MUGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	800
CGTGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	840
CGTGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	880
CGTGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	920
CGTGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	960
CGTGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	1000
ATCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	1040
CGTGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	1080
ATCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	1120
TGCTAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	1160
TGCTAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	1200
CGTGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	1240
CGTGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	1280
ACGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	1320
CGTGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	1360
CGTAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	1400
CGTGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	1440
ATCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	1480
TGCTAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	1520
ACGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	1560
CGTGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	1600
CAATTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	1640
CGTGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	1680
CGTGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	1720
TGCTAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	1760
CGTGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	1800
TGCTAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	1840
ATCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	1880
CGTGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	1920
CGTGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	1960
TGCTAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	2000

GEORGE 2007

Figure 24B

nsdyvggahitvrladlatmpvrladitvivrugantgil 40  
amnskasigtvifcraaryvdrvlikgicqglyxatara 80  
tannyaavlaangvuggdvvgiamlmspsvllanlatvka 120  
galagmlyirgrgevlahsagilicakvliassdlyvavae 160  
cgaagrvagowltveavvafatagatogasasavvaki 200  
tatyitfuggtugpkasmmimwlravviggagilrlikg 240  
sdilyvsciplyimalltwavsvinegatlalgcetiaer 280  
fmderlanatavwvigeicryllnqaigpuraingvvi 320  
cagclpgeiwdefmrigvarvcefyvasegmsafini fm 360  
vpartagvsqumlatfvevtilcgcclixdasgmvrvpdeq 400  
gilisavvrlgndgvtogvaseidivmafrdgcdwnt 440  
gavmspamgzaafvurigdrfwagewathgveaalas 480  
cgtveactvvgvcpitggagagaaatiragastoggaia 520  
rtvychlpgyalpifrvvgslahittuikavalmogay 560  
cathicpilyvlagpdeagvvyvayvasevligmpog 597

Figure 25

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	8												

Figure 26



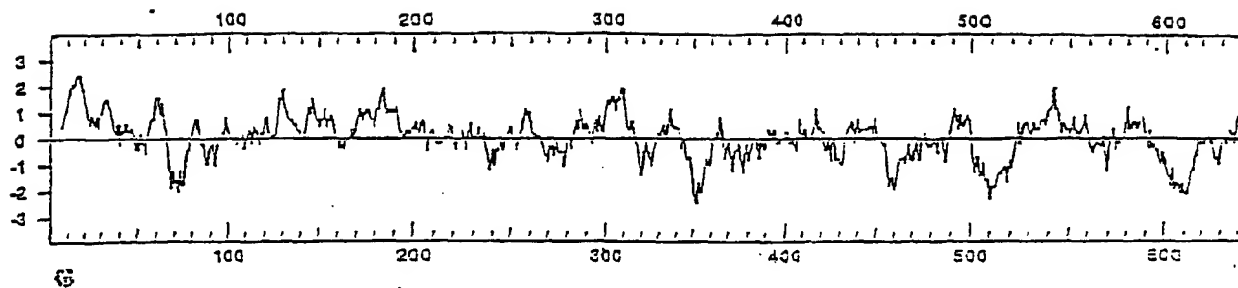


Figure 28A

			n	D(%)	MS	MS(%)
A	ala	alanine	44	3.3	4546	6.4
C	cys	cysteine	12	2.1	1547	2.2
D	asp	aspartic acid	30	4.3	1420	4.3
E	glu	glutamic acid	11	4.3	4000	4.4
F	pro	phenylalanine	23	4.3	4564	4.3
G	gly	glycine	43	9.3	1332	3.1
H	his	histidine	11	2.3	1781	2.3
I	ile	isoleucine	23	4.3	1379	4.3
K	lys	lysine	21	3.4	2613	4.3
L	leu	leucine	27	11.3	8707	12.3
M	met	methionine	11	1.7	1441	2.3
N	asn	asparagine	12	2.3	1710	2.4
P	pro	proline	23	4.3	2414	4.3
Q	gln	glutamine	25	1.3	1201	4.3
R	arg	arginine	43	7.3	7848	10.3
S	ser	serine	11	5.1	2572	4.3
T	thr	threonine	27	4.3	2713	3.3
V	val	valine	21	7.3	2023	7.1
W	trp	tryptophan	9	1.4	1574	2.4
X	tyr	tyrosine	-	-	-	-
Y	tyr	tyrosine	14	1.7	1911	3.3
Z	-	-	-	-	-	-

Figure 28B

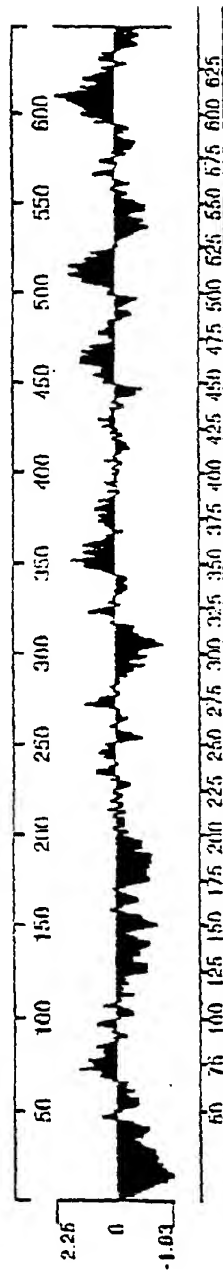


Figure 28C



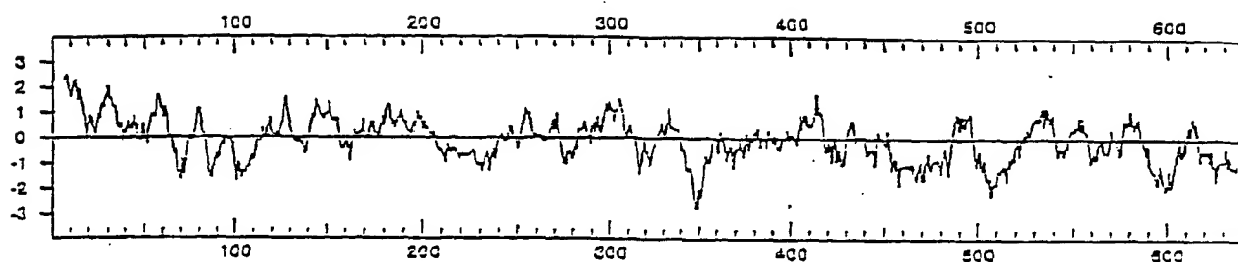


Figure 29A

		n	n(%)	SN	SN(%)	
A	ala	alanine	46	7.1	3267	4.5
C	cy	cytosine	14	2.1	1546	2.1
D	asp	aspartic acid	32	5.1	3795	5.1
E	glu	glutamic acid	32	5.1	4233	5.9
F	phe	phenylalanine	34	5.1	5000	6.9
G	gly	glycine	54	8.1	1079	4.1
H	his	histidine	12	1.9	1544	2.1
I	ile	isoleucine	30	4.7	3192	4.7
K	lys	lysine	31	4.8	3970	5.5
L	leu	leucine	76	12.3	8354	11.9
M	met	methionine	12	1.9	1571	2.1
N	asn	asparagine	21	3.1	2194	3.1
P	pro	proline	11	4.3	2008	4.2
Q	gln	glutamine	23	3.5	2945	4.1
R	arg	arginine	45	7.1	7024	9.3
S	ser	serine	19	2.4	1046	4.2
T	thr	threonine	12	5.0	3222	4.2
V	val	valine	46	7.1	4527	5.1
W	trp	tryptophan	8	1.2	1488	2.1
X	unk	unknown	-	-	-	-
Y	cy	cytosine	14	1.9	4076	5.7
Z	---	---	-	-	-	-

Figure 29B

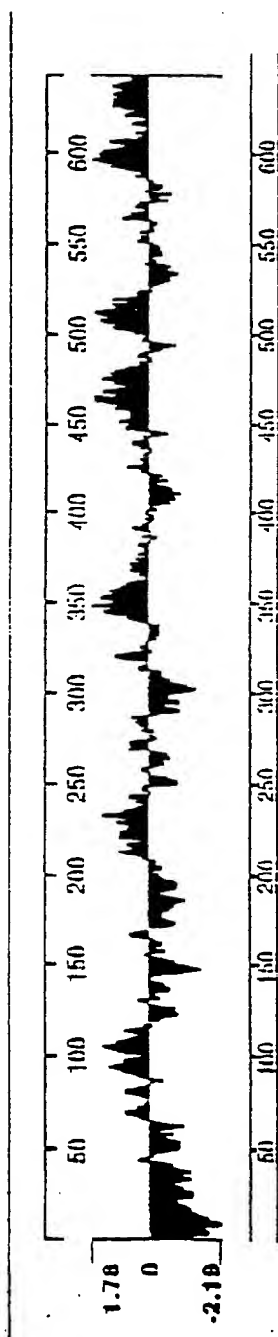
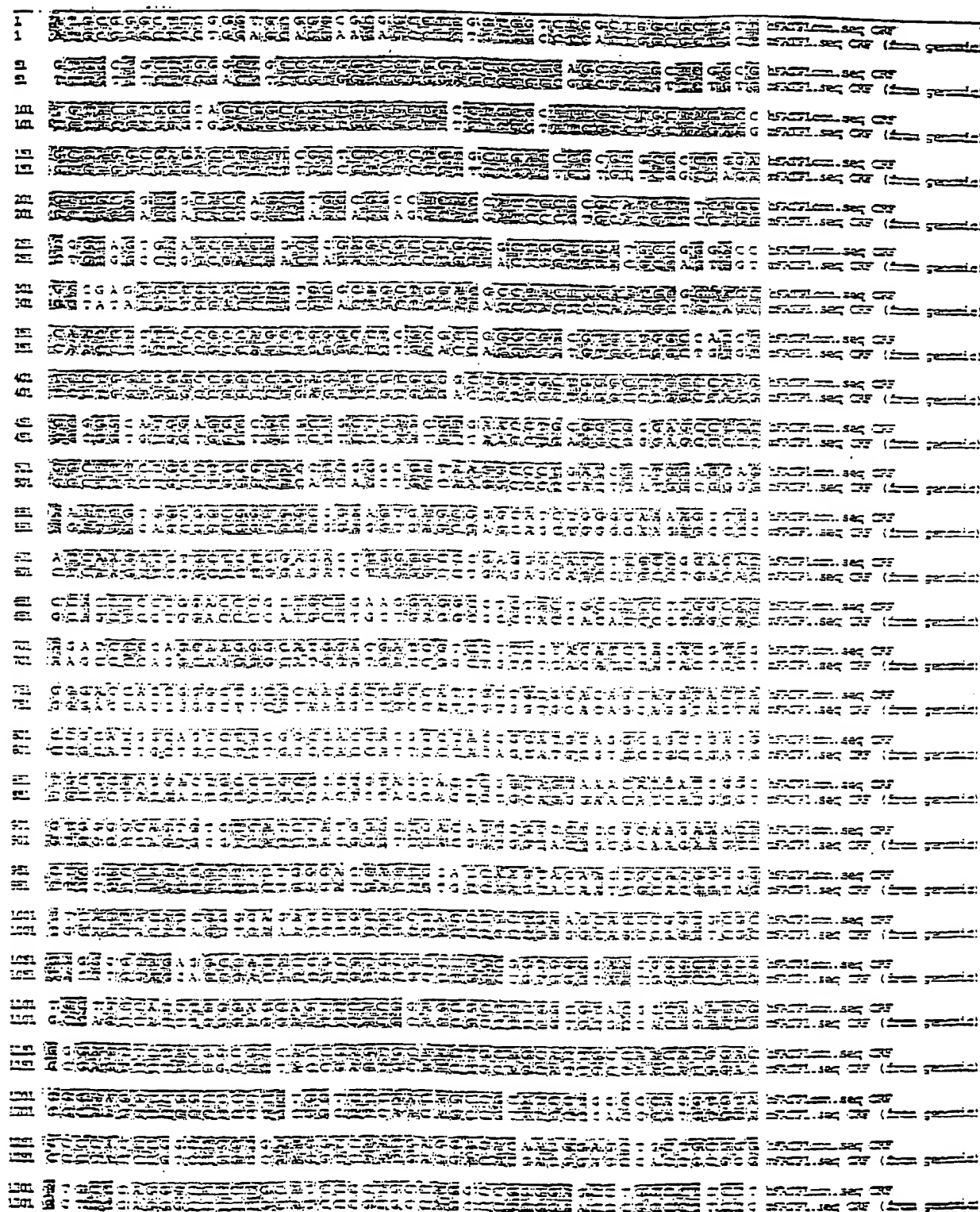
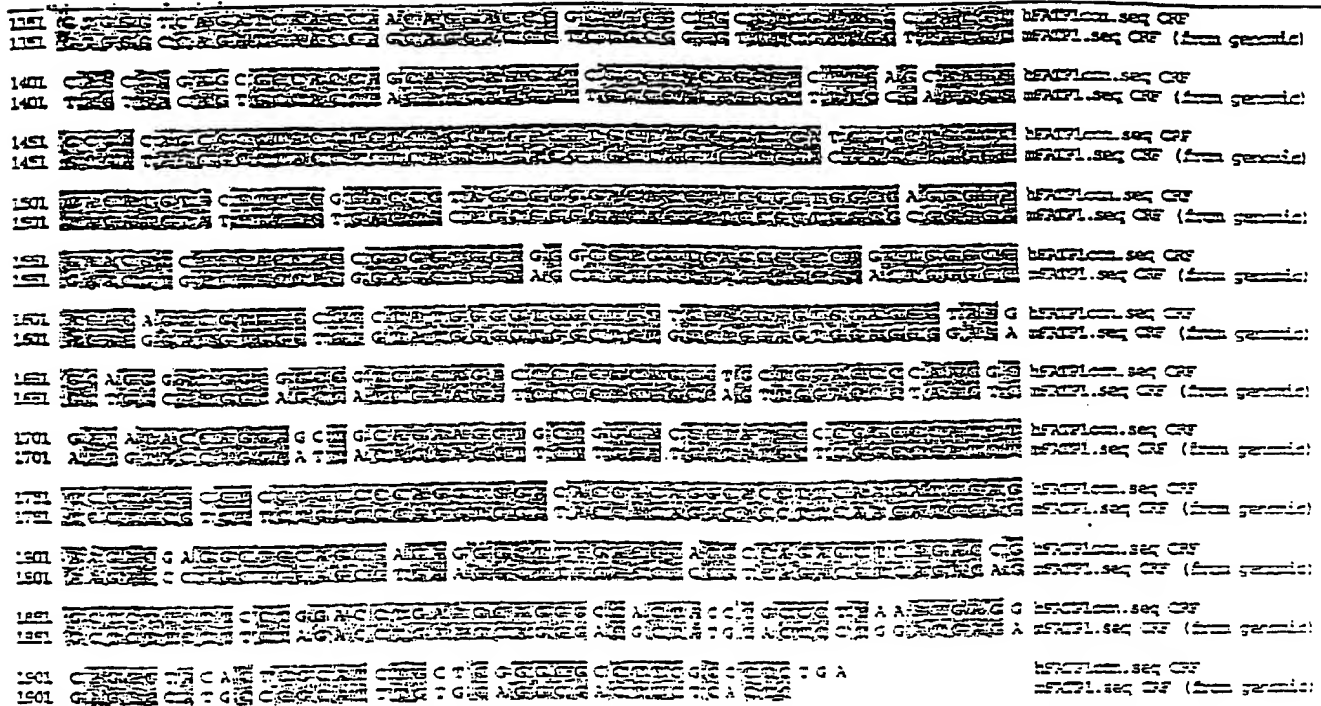


Figure 29C





Definition: 'Consensus #1': Peaks (with solid high yellow) indicates that match the consensus signal 'Consensus #1' exactly.

FIG. 30B

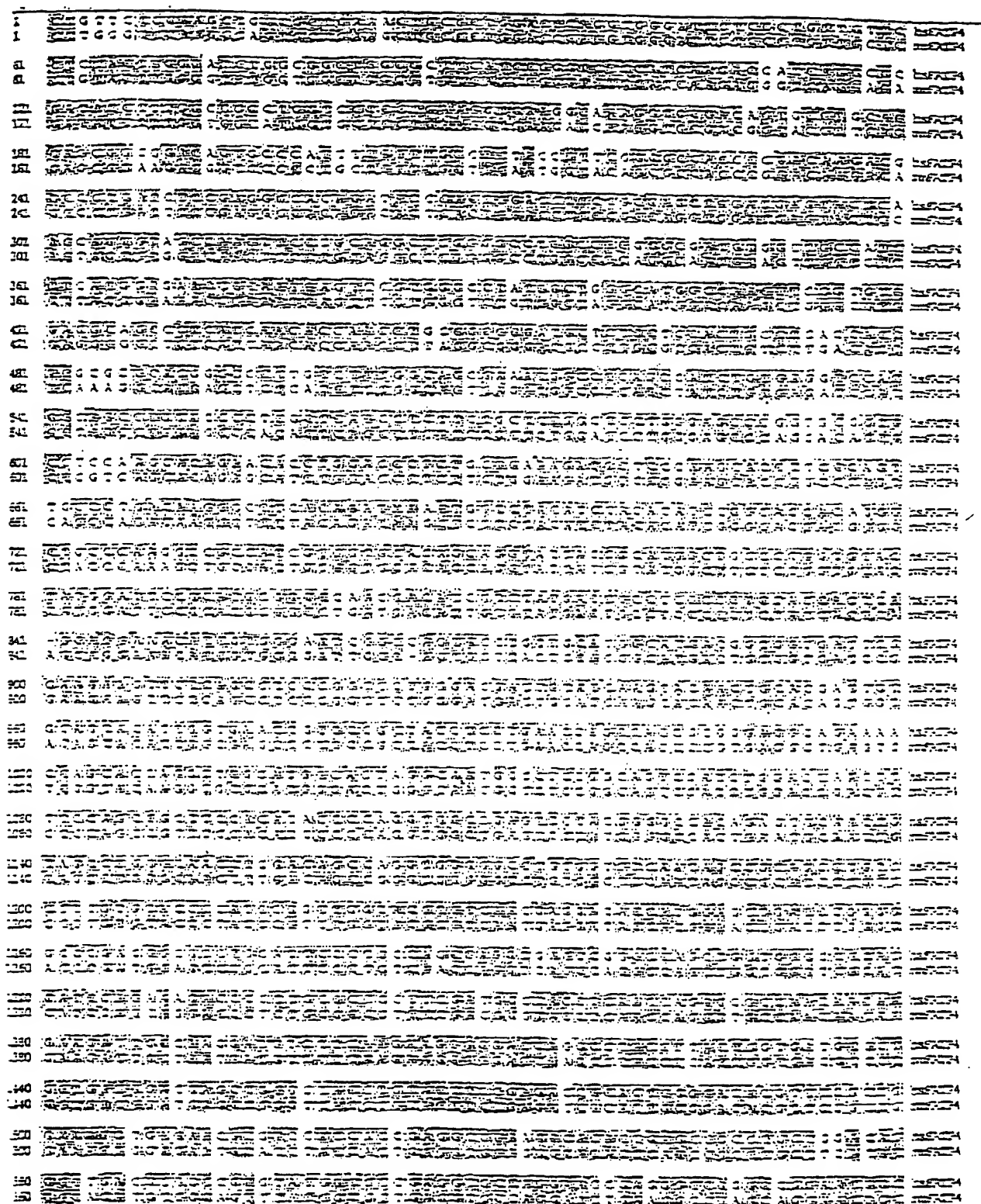


Figure 3/A

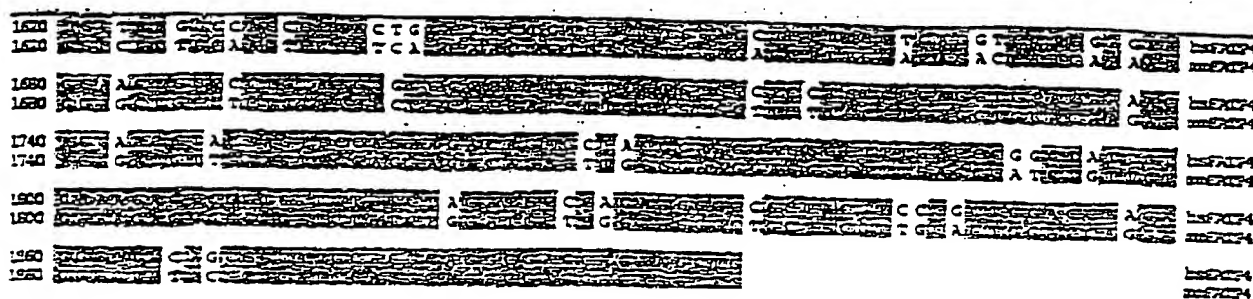


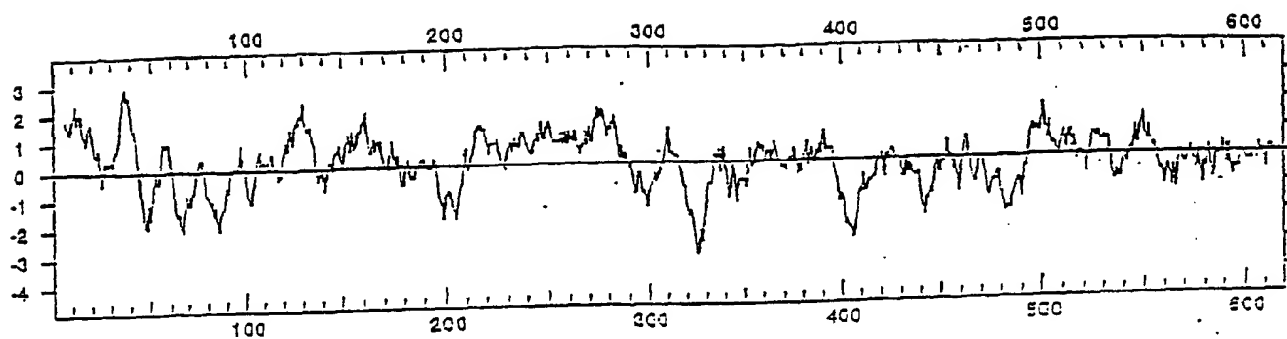
Figure 31B











			2	2(3)	2(4)	2(5)
2	2(1)	2(2)	12	2.1	2144	2.1
3	3(1)	3(2)	14	2.1	1442	2.1
4	4(1)	4(2)	34	2.1	3412	2.1
5	5(1)	5(2)	11	2.1	1100	2.1
6	6(1)	6(2)	14	2.1	1400	2.1
7	7(1)	7(2)	44	7.1	4400	7.1
8	8(1)	8(2)	12	2.1	1200	2.1
9	9(1)	9(2)	37	6.1	3724	6.1
10	10(1)	10(2)	46	7.1	4648	7.1
11	11(1)	11(2)	77	12.1	7747	12.1
12	12(1)	12(2)	11	1.1	1144	1.1
13	13(1)	13(2)	21	1.4	2194	1.4
14	14(1)	14(2)	21	1.4	2109	1.4
15	15(1)	15(2)	13	2.1	1305	2.1
16	16(1)	16(2)	27	4.4	2714	4.4
17	17(1)	17(2)	40	5.1	4017	5.1
18	18(1)	18(2)	30	4.1	3012	4.1
19	19(1)	19(2)	51	8.1	5052	8.1
20	20(1)	20(2)	11	1.1	1046	1.1
21	21(1)	21(2)	-	-	-	-
22	22(1)	22(2)	25	4.2	4209	4.2
23	23(1)	23(2)	-	-	-	-

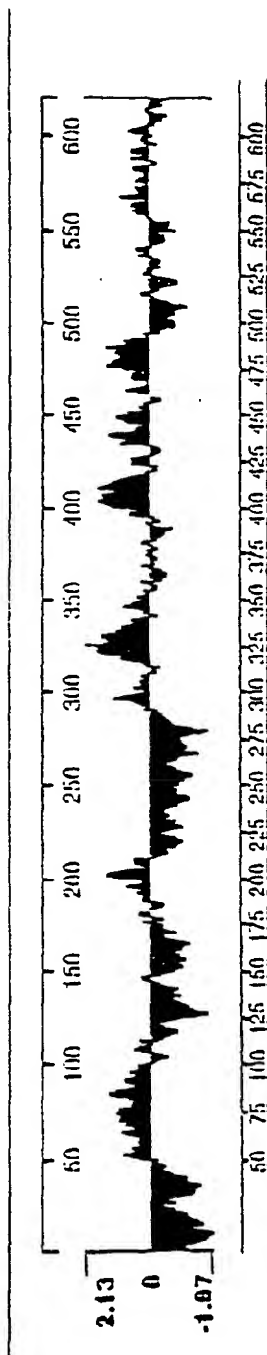


Figure 35C

[illegible]

Figure 36

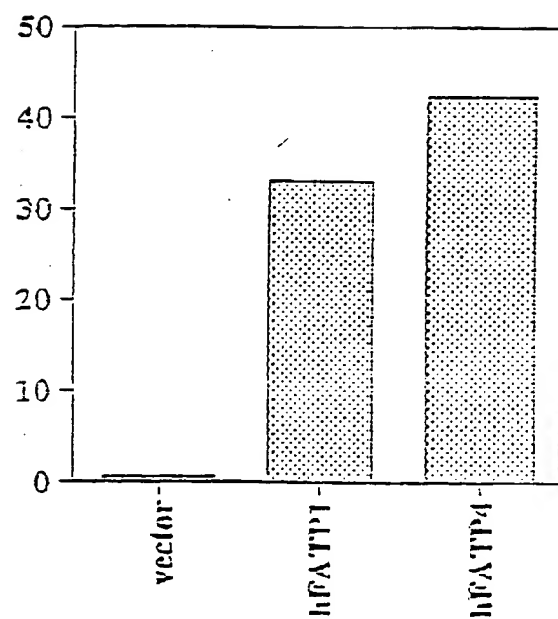


Figure 37

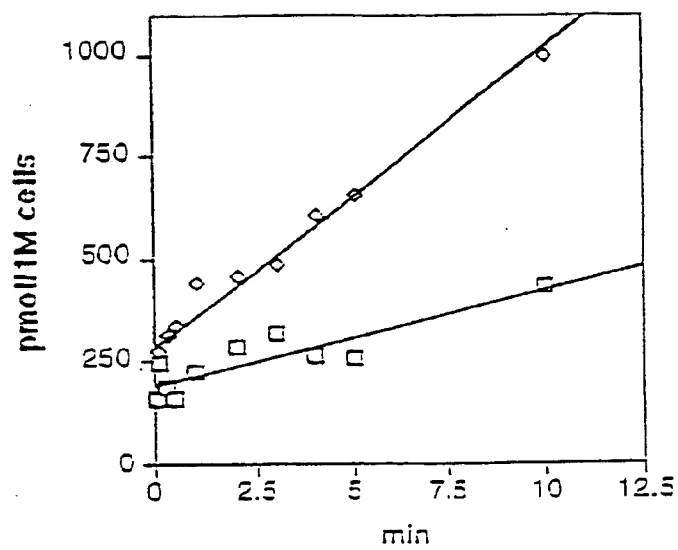


Fig. 38

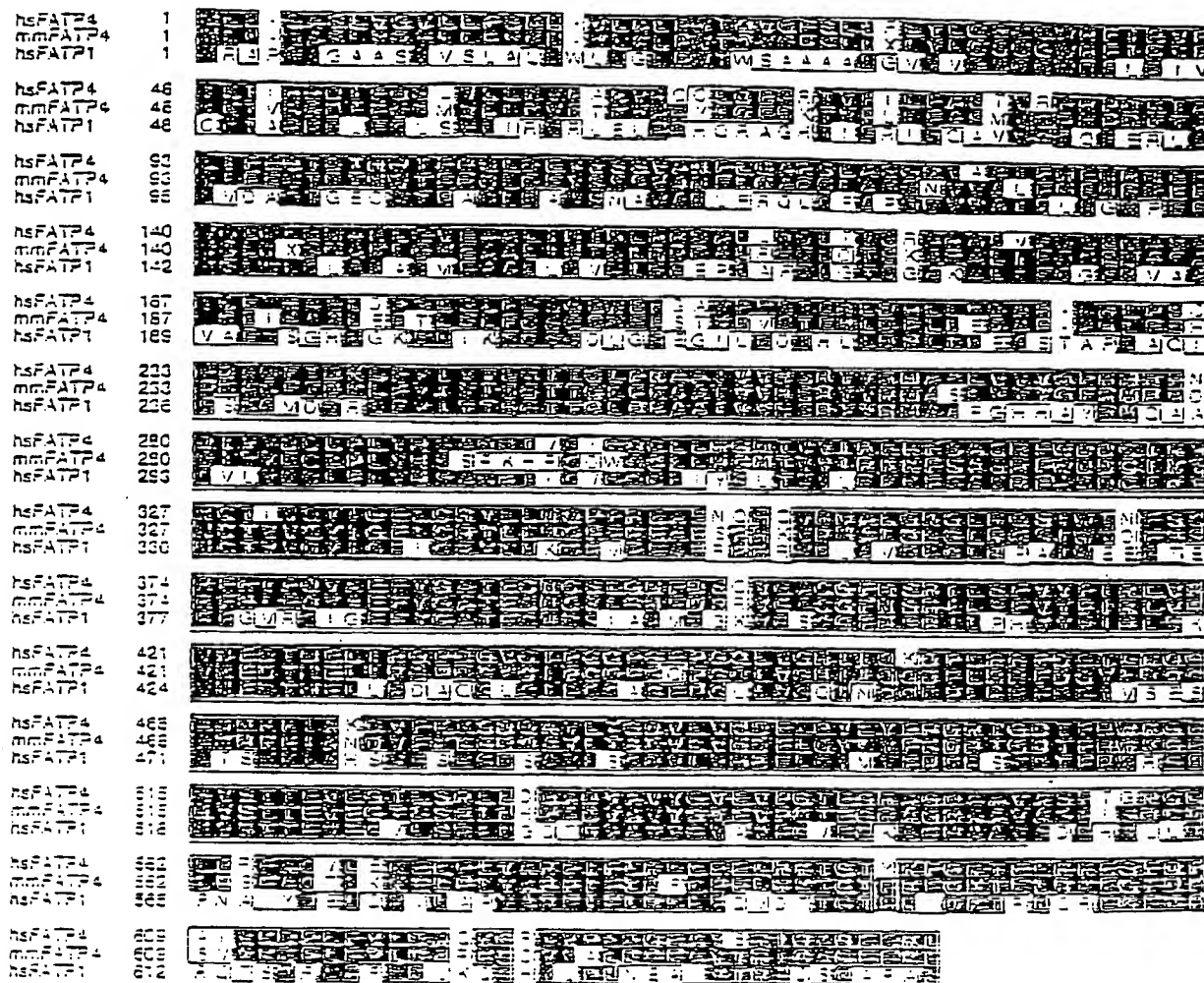


Fig 39

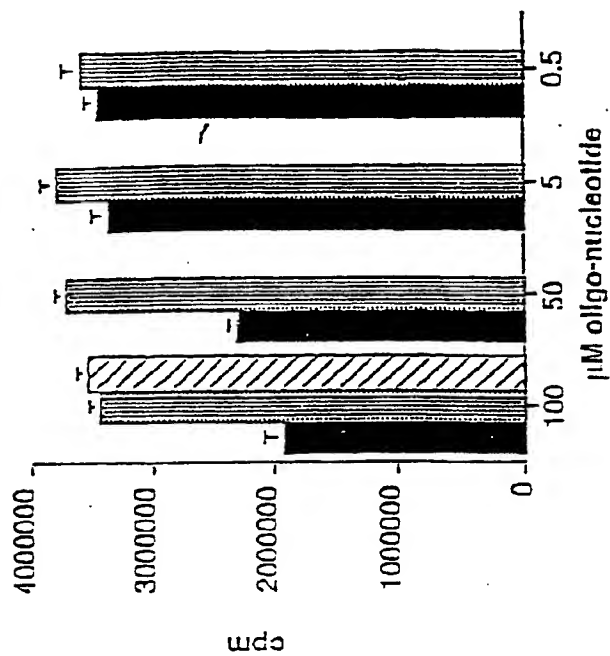


Fig. 41

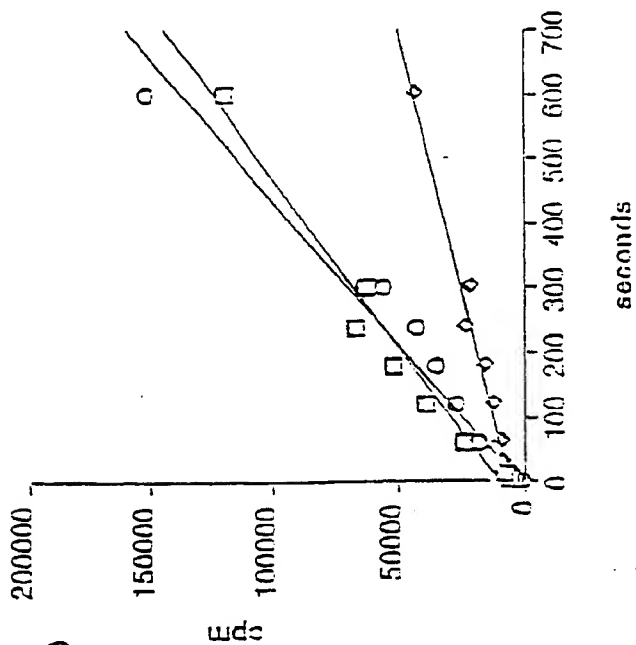


Fig. 40

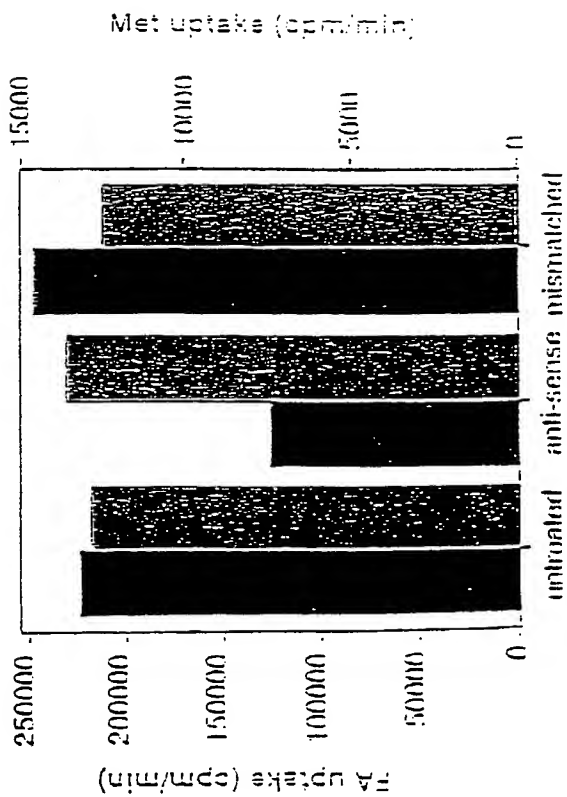


Fig. 42



1. The first step in the process is to identify the problem or issue that needs to be addressed. This involves gathering information and understanding the context of the problem.

Figure 43A

[illegible]

Figure 43B

10 20 30 40  
TCGACCCACGGGCTCCGGGACCCCAAAGCAGAAGCCCCCA 40  
CAGTAGGCACAGCGCACCCCAAGAAGGGTCCAGGAGTCTGC 80  
AGAAACAGAAAGGTCCCCGGCCTCAGCCTCCTAGTCCCTG 120  
CCTGCCCTCCTGCCTGAGCTTCTGGGAGACTGAAGGCACGG 160  
CTTGACAGCTTCAGGAATGCGGGCTCCGGGTGCGGGCGCGGC 200  
210 220 230 240  
CTCGGTGGTCTCGCTGGCGCTGTTGTGGCTGCTGGGGCTG 240  
CCGTGGACCTGGAGCGCGGCAGCGCGCTCGGGCTGTACG 280  
TGGGCAGCGCGGCTGGCGCTTCTGCGCATCGTCTGCAA 320  
GACCGCGAGGGGAGACCTCTTGGGTCTCTCTGTGCTGATC 360  
CGCGTGCCTTGGAGCTGCGGGCGGCACCAAGCTGCCGGGC 400  
410 420 430 440  
ACACCATCCCGGCACTCTTTCAGGGCGTAGTGCAGCGACA 440  
GCCCCAGCGCCTGGCGCTGGTGGATGCCGGGACCGGGAG 480  
TGCTGGACCTTTGGGAGCTGGAGCGCTACTCCAAATGCCG 520  
TAGCCAACTCTTGGCGCAGCTGGGCTTCGGCGCGGGCGA 560  
CGTGGTGGCCATCTTCTCGAGGGCGGGCGGAGTTCTGCTG 600  
610 620 630 640  
GGGCTGTGGCTGGGCTGGGCAAGGGGGGATGAGGGCG 640  
CGCTGCTCAAGCTGAACCTTGGCGCGGAGCGGCTGGGCTT 680  
CTGGCTGGGCACTTGGGGCGCTAAGGGCTTGAATTTTGG 720  
GGAGAAATGGTGGGGGCGGTGGCGGAAATGAGCGGGCATC 760  
TGGGGAAGAGTTGATCAGATCTGCTCTGGAATCTTGGG 800  
810 820 830 840  
GCCCAGGGGATCTTGGCGGACACCCACCTGCTGGACCGG 840  
CTGCTGAAGGAGGCTCTACTGCCCCCTTGGGACAGATCC 880  
CCAGCAAGGGGATGGAGCGATCGTCTTTCTTACATCTACAC 920  
GTGCGGGACCAAGGGCTGCCCCAAGGCTGCCATTGTCTGT 960  
CACAGCAGGTACTACCGCATGGCAGCTTCCGGCAACAGG 1000  
1010 1020 1030 1040  
CCTACCGCATGCAGGGCGGTGACCTGCTCTATGACTGCCT 1040  
GCCCCGTGTACCACTCGGCAGGAAACATCATCGGGCTGGGG 1080  
CAGTGTCTCATCTATGGGCTGACAGTCTCTCTCGGCAAGA 1120  
AATTCTCGGGCAGCGCTTCTGGGACGACTGCATCAAGTA 1160  
CAACTGCACGGTGGTTTCAGTACATCGGGGAGATCTGCCGC 1200

Fig. 44A

1210	1220	1230	1240
TACCTGCTGAAGCAGCCGGTGC	CGGAGGCGGAGAGGCGAC	1240	
ACCGCGTGCGCTGGCGGTGGG	AACGGGCTGCGTCCTGC	1250	
CATCTGGGAGGAGTTACCGAG	CGCTTCGGCGTACGCCAA	1320	
ATCGGGGAGTTCTACGGCGCC	ACCGAGTGCAACTGCAGCA	1350	
TTGCCAACATGGACGGCAAGG	TGGCTCCTGTGGTTTCAA	1400	

1410	1420	1430	1440
CAGCGCATCCTGCCCCACGTGT	ACCCCATCCGGCTGGTG	1440	
AAGGTCAAATGAGGACACAAT	TGGAGCTGCTGGGGATGCCC	1450	
AGGGCCTCTGCATCCCCCTGC	CAGGCCGGGAGCCTGGCCT	1520	
CCTTGTGGGTGAGATCAACCA	ACAGGACCGCTGGCGCGC	1550	
TTCCATGGCTATGTACGGAG	AGCGCCACCAGCAAGAAGA	1600	

1610	1620	1630	1640
TGCCCCACAGCGTCTTTCAGCA	AGGGCGACAGCGCTACCT	1640	
CTCAGGTGACGTGCTAGTGAT	TGGATGAGCTGGGCTACATG	1650	
TACTTCCGGGACCGTAGCGGG	GACACCTTCGGCTGGCGAG	1720	
GGGAGAACGTCTCCACCACCG	AGGTGGAGGGCGTGGTGA	1750	
CCGGCTGCTGGGGCCAGACAG	AGGTGGCGGTCTATGGGGTG	1800	

1810	1820	1830	1840
GCTGTTCCAGGAGTGGAGGGT	AAGGCAGGGATGGCGGCG	1840	
TGGCAGACCCCCACAGCCTGT	CTGGACCCCAACGGGATATA	1850	
CCAGGAGCTGCAGAAAGGTGT	GGCACCTATGGCGGGCCC	1920	
ATCTTCCCTGGCGCTCTCTGC	CCCCAGGTGGACACCACAGCA	1950	
CCTTCAAGATCCAGAAAGAG	AGGCTGGCAGCGATAGGGCTT	2000	

2010	2020	2030	2040
TGACCCAGGGCAGACCTCAG	AGCGGCTCTTCTCTCTCTGG	2040	
CTGAAGCAGGGCCACTACCT	GGCCTTAAATGAGGCAGTCT	2050	
ACACTGGCATCTGCTCGGG	CGGCTTCGGCTCTGAAGCTG	2120	
TTCTCTACTGGCCACAACT	CTGGGGCTGGTGGGAGAGG	2150	
CCAGCTTGAGCCAGACAGCG	CTGCCAGGGGTGGCGGCT	2200	

2210	2220	2230	2240
AGTACACACCEACCTGGCCG	AGCTGTACCTGGCACGGCC	2240	
ATCCTGGACTGAGAACTGG	AACTCAGAGGAACCGCTGC	2250	
CTCTCTGCTGCTTGGTGGC	CTGTGTCTGGCTCTCTCTC	2320	
CTGCTTTTCAGCCTCTGCT	CTCTCTCCATCTCTCTCTCT	2350	
CTGGCTTAACTCTCTCTCT	CTCTCTCTCTCTCTCTCT	2400	

2410	2420	2430	2440
TTCTTTTTTTTTTAAGATAG	AGTCTCACTCTGCTGCCCC	2440	
CTAGAGTGCAGTGGTGGGA	TCTCGGCTCACTGCAACCT	2450	
GCCTCTGGGGTTCAAGTG	ATCTCTCCACCTCAGCCTCT	2520	
GAGTAGCTGGGATTACAGG	CACCGCCACCACGTCCAGCT	2550	
AATTTTTATATTTTACTAG	AGACGGGGTTTCACCATGTT	2600	

Fig. 44B

2610	2620	2630	2640
GGTCAGGCTGGTCTTGAACCTCCTGACCTCAGGTGATCCGC	2640		
TGGCCTCGGCCTCCACAGAGTGCTGGGATTATAGGCGTGAG	2680		
CCTCTGGCCCCGGCCTTTCTTTTTCTCTCTCTCTCTCTCT	2720		
GAGAGTGGAAACACACGTGTCTTGGGAGCTGCATCTTGTGT	2760		
AGGGTCCAGCTGCTTTTGGGGACTGCAGGAATCATCTCCC	2800		

2810	2820	2830	2840
CTGGGCCCTGGACTCGGACTGGGGCCTCCCCACCTCCCTC	2840		
TGGGCTGTGCCTTACGGAGCCCCAATCCAGGCCTCCTGTG	2880		
GCTGTTGGGTTCCAGATGCTGCAGCTCCATGTGACTTCCA	2920		
AGCAGGCCCCCTCGGCCCTCCCCTGCTGAATGGAGGAGCCGG	2960		
GGTCCCCCAGGCCAAGTGGAAAACTCCCAGGCTAGGCCA	3000		

3010	3020	3030	3040
ATTGCGCTTTTGCACCTTCCCCGTTCTGTACATTTCCCCA	3040		
GGCCCCACCTTCCCCCTCCTGATGCCCTGAAAGCTTCCGGA	3080		
TTGACTGTGACCACTTGGATGTCAACCACTGTGAGCCCCCT	3120		
CCTTGATGTCCCCATTTAGCCATCTCCATGGAGCTCCTGC	3160		
TGGAGGGGCCCTGAACCTTGCACCTGCGTGGCTGCCAGGCCA	3200		

3210	3220	3230	3240
GCTGCCTCCTGTCTCTGGGAGGAGGCCTCCTGGGTGTCTC	3240		
ATCTGGTGTGTCTACTGGAGGGTCCCACAGGAGAGGCAGC	3280		
AGAGGGGTGAGGGGAGGTCTCTTCCGGGGGTTGGCCTCT	3320		
CAAGCCTCAGGGGTTCTAGCCTGTTGAATATACCCCACCT	3360		
GGTGGGTGGCCCCCTCCGATGTCCCCACTGATGGCTGTGAC	3400		

3410	3420	3430	3440
ACCGTGTGGTGGCGATGTCCCAGACAAATCCACCAAGAC	3440		
GGCCCCAGACATCCCTACTGGCTTCCCTGGTGGCTCATCTC	3480		
GAACATCCACGCCAAGCCTTTCTGGGGCCGGCCACCCAGGC	3520		
CGCCTGTCCGTCTGTCTCTCCCTCCAGCAGCAACCCCTGGC	3560		
CCCTGGAGTGGTGGGGCCATGGCAAGAGACACCGTGGCGT	3600		

3610	3620	3630	3640
CTCATGTGAACCTTTCTGGGCACTGTGGTTTTTATTTCTTA	3640		
ATTGATTTAAGAAATAAACCTGAAGACCGTCTGGTGAATA	3680		
AAAAAAAAAAAAAAAA	3694		

Fig. 44C

10	20	30	40
MRAPGAGAASVVSALLWLLGLPWTWSAAAAALGVYVGS	40		
WRFLRIVCKTARROLFGLSVLIRVLELRRHCRAGHTIPR	80		
IFQAVVGRCFERLALVQAGTGECWTFAGLDAYSNAVANLF	120		
RGUGFAPGDVVAIFLEGRPEFVGLWLGLAKAGMEAALLNY	160		
NLRREPLAFCLGTSGAKALIFGGEMVAAVAESVCHLCKSL	200		
210	220	230	240
IKFCSGGLGPEGILFOTHLLDPLLKEASTAPLAGIPSKGM	240		
QORLPYIYTEGTTGLPKAAIVVHSRYRMAAFGRHAYRMC	280		
AAOVLYDCLPLYHSAGNIIGVGOCLLYGLTVVLRKKFSAS	320		
RFWCCGCKYNCTVVGYYIGEICRYLLKGFVREARRHVRU	360		
AVGNGLRPAIWEEFTERFGVRCIGEFYGATECNGSIANMD	400		
410	420	430	440
GKVGSGGFNSRILPHVYFIRLVKYNEDTHELLRCAGGLOI	440		
FOGAGEPGLLVGGINGCDPLRRFDGYVSESATSKXIAHSV	480		
FSKGDSAYLSSGOVLVMDELGYMKFRDRSGCTFRWRGENYS	520		
TTEVEGVLSRLLGGTGAVVYGVAVFGVEGKAGMAAYADPH	560		
SLLOPNAIYGELOKVLAPYARRIFLRLLPGVOTTGTGTXIC	600		
610	620	630	640
KTRLQREGFDPROTSCRLFFLOLKGGHYLPLNEAVYTRIC	640		
SGAFAL	646		

Fig. 45

```

      10      20      30      40
.....
GGAATTCCAAAAAAAAAATACGACTACACCTGCTCCGG 40
AGCCCCGCGGGGTACCTGCAGCGGAGGAGCTCTGTCTTCC 60
CCTTCATCTCAGCGGAGCCCCGGCGTCCCGCCGCGTGCGCC 120
CCGGCAGCCCCGCCAGTCCGCCCGGAGCCCCGCCAGTCG 160
CCGCGCTGCACGCCCGGGGTGAACCTCTGCCCTCGCTGG 200

      210      220      230      240
.....
GACAGAGGGCCCCCGCAGCCGTCATGCTTTCCGCCATCTAC 240
ACAGTCCTGGCGGGAAGTGTGTTTCTGCGGCTCTGGTGA 280
ACCTCTGCTGCCCATACTTCTTCCAGGACATAGGCTACTT 320
CTTGAAGGTGGCCCGCGTGGGCGGGAGGGTGGCCAGCTAC 360
GGGCAAGGGCGGCGGCGCGCACCATCTGCGGGCGTTC 400

      410      420      430      440
.....
TGGAGAAAGCCCGCCAGACGCCACACAAGCCTTTTCTGCT 440
CTTCCGGGACGAGACTCTCACCTAGCGCGAGGTGGACCGG 480
CGCAGCAATCAAGTGGCCCCGGCGCTGCAGGACCACCTCG 520
GCCTGCGCCAGGGAGACTGCGTGGCGCTCTTATGGGTAA 560
CGAGCGGGCTACGTTGTGGCTGTGCTGGGGCTGGTGAAG 600

      610      620      630      640
.....
CTGGGCTGTGCCATGGCGTGGCTCAATTACAAATCCGGG 640
CGAAGTCCCTGCTGCACTGCTTCCAGTGTGGGGGGGAA 680
GGTGTGCTGGTGTGCGCCAGAACTACAAAGCAGTGTGAA 720
GAGATACTGCCAAAGCCTTAAAAAAGATGATGTGTCCATCT 760
ATTATGTGAGCCAGAACTTCTAACACAGATGGGATTGACTC 800

      810      820      830      840
.....
TTTCTGGACAAAGTGGATGAAGTATCAACTGAACCTATC 840
CCAGAGTCAATGGAGGTGTGAAGTCACTTTTCCACTCCTG 880
CCTTATACATTTTACTTTCTGGAACCAAGGCTTTCCAAA 920
AGCAGCCATGATCACTCATCAGCGCATATGGTATGGAACT 960
GGCTCACTTTTGTAAAGCGGATTGAAGGCAGATGATGTCA 1000

      1010      1020      1030      1040
.....
TCTATATCACTCTGCCCTTTTACCACAGTGTGCACTACT 1040
GATTGGCATTACCGGATGTATTGTGGCTGGTCTACTCTT 1080
GCCTTGGCGACTAAATTTTCAGCCAGCCAGTTTGGGATG 1120
ACTGCAGAAAATACAACGTCAGTGTCAATCAGTATATCG 1160
TGAAGTCTTGGGTATTTATGCAACTCACCACAGAAACCA 1200

```

Fig. 46 A

1210	1220	1230	1240
AATGACCGTGATCATAAAGTGAGACTGGCACTGGGAAATG	1240		
GCTTACGAGGAGATGTGTGGAGACAATTTGTCAAGAGATT	1280		
TGGGGACATATGCATCTATGAGTTCTATGCTGCCACTGAA	1320		
GGCAATATTGGATTTATGAATTATGCCGAGAAAAAGTTGGTG	1360		
CTGTTGGAAGAGTAAACTACCTACAGAAAAAATCATAAC	1400		
1410	1420	1430	1440
TTATGACCTGATTAATATGATGTGGAGAAAGATGAACCT	1440		
GTCCGAGATGAAAAATGGATATTGCGTCAGAGTTCCCAAAG	1480		
GTGAAGTTGGACTTCTGGTTTGCAAAATCACACAACCTTAC	1520		
ACCATTTAATGGCTATGCTGGAGCAAAGGCTCAGACAGAG	1560		
AAGAAAAAATGAGAGATGTCTTAAGAAAGGAGACCTCT	1600		
1610	1620	1630	1640
ATTTCAACAGTGGAGATCTCTTAATGTTGACCATGAAAA	1640		
TTTCATCTATTTCCACGACAGAGTTGGAGATACATTCGGG	1680		
TGGAAAGGGGAAAAATGTGGCCACCACTGAAGTTGCTGATA	1720		
CAGTTGGACTGGTTGATTTTGTCCAAAGAAATGTTTA	1760		
TGGAGTGCATGTGCCAGATCATGAGGGTGGCATTGGCATG	1800		
1810	1820	1830	1840
GCCTCCATCAAAATGAAAGAAAAACCATGAATTTGATGGAA	1840		
AGAAACTCTTTCCAGCACATTGCTGATTACCTACCTAGTTA	1880		
TGCAAGGGCCCCGGTTTCTAAGAATACAGGACACCATTGAG	1920		
ATCACTGGAACTTTTAAACACCGCAAAATGACCCCTGGTGG	1960		
AGGAGGGGCTTTAACCCCTGCTGTTCATCAAAAGATGGCTTGT	2000		
2010	2020	2030	2040
TTTCTTGGATGACACAGCAAAAAATGTATGTGCTATGACT	2040		
GAGGACATCTATAATGGCATAAGTGTCTAAAACCTGAAAC	2080		
TCTGAATATTCCAGGAGGATAAATCAACATTTCCAGAAA	2120		
GAAACTGAATGGACAGGCACTTGATATAATCCAACTTTAA	2160		
TTTGATTGAAGATTGTGAGGAAAAATTTGTAGGAAAATTTGC	2200		
2210	2220	2230	2240
ATACCCGTAAAGGGAGACTTTTTTAAATAACAGTTGAGTC	2240		
TTTGCAAGTAATAAGATTTAGAGATTATTTTTCAGTG	2280		
TGCACCTACTGTTTGTATTTGCCAAACTGAGCTTGTGGAG	2320		
GGAAAGGCATTATTTTTTAAAAATACCTAGTAATTTAAATGA	2360		
AC	2362		

Fig. 4B

```
      10      20      30      40
.....-
MLSAIYTVLAGLLFLPLLVLNLCOPYFFGDIQYFLKVAAVG 40
RRVRSYGGRRFPARTILRAFLEKARCTPHKFFLLFRGETLT 80
YAGVORRSNGVARALHGHGLGLRGGOCVALLMGNEFAYVWL 120
WLGLVKLGCAHACLNYNIKAKSLLHCFGCCGAKVLLVSFE 160
LGAAVEEILFSLKKQDVSIYYVSRTSNTDGIQSPFCKVDE 200
      210      220      230      240
.....-
VSTEPIPESWRSEVTFSTPALYIYTSGETTGLPKAAMITHG 240
RIWYGTGLTFVSGLKADQVIYITLFFYHSAALLIGIHCCI 280
VAGATLALRTKFSASQFWDDCRKYNVTVIQYIGELLRYLC 320
NSPCKFNDRGHKVRLALGNGLRGQVWRQFVKRFQDICIYE 360
FYAATEGNIQFMNYARKVGAVGRVNYLGKKIITYGLIKYS 400
      410      420      430      440
.....-
VEKDEPYRDENGYQVRVPKGEVGLLVCKITQLTFENGAYG 440
AKACTEKKKLROVFKKGOLYFNSGOLLMVCHENFIYFHR 480
VGGTFRWKGENVATTEVAGTVGLVDFVGEVNVYGVHYFCH 520
EGRIQMASIKMKENHSEFGKKLFQHIADYLPSYARPRFLR 560
IQQTIEITSTFKHRKNTLYEEGFNPAVIKDLYFLDGTAK 600
      610      620      630      640
.....-
MYVPMTEDLYNAISAKTLKL 620
```

Fig. 47



10 20 30 40  
AAGTTCTCGGCTGGTCAGTTC TGGGAAGATTGCCAGCAGC 40  
ACAGGGTGACGGTGTTCCAGTACATTGGGGAGCTGTGCCG 80  
ATACCTTGTCAACCAGCCCCCGAGCAAGGCAGAACGTGGC 120  
CATAAGGTCCGGCTGGCAGTGGGCAGCGGGCTGGCCCCAG 160  
ATACCTGGGAGCGTFTTGTGCGGGCGCTTCGGGCCCCCTGCA 200  
210 220 230 240  
GGTGCTGGAGACATATGGACTGACAGAGGGCAACGTGGCC 240  
ACCATCAACTACACAGGACAGCGGGGGCGCTGTGGGGCGTG 260  
CTTCCTGGCTTTACAAGCATATCTTCCCTTCTCCTTGAT 320  
TCGCTATGATGTCAACCACAGGAGAGCCAAATTCGGGACCCC 360  
CAGGGGCACTGTATGGCCACATCTCCAGGTGAGCCAGGGG 400  
410 420 430 440  
TGCTGGTGGCCCCGGTAAGCCAGCAGTCCCCATTCTCTGG 440  
CTATGCTGGCGGGCCAGAGCTGGCCCCAGGGGAAGTTGCTA 480  
AAGGATGTCTTCCGGCCTGGGGATGTTTTCTTCAACACTG 520  
GGGACCTGCTGGTCTGCGATGACCAAGGTTTTCTCCGCTT 560  
CCATGATCGTACTGGAGACACCTTCAGGTGGAAGGGGGAG 600  
610 620 630 640  
AATGTGGCCACAACCGAGGGTGGCAGAGGTCTTCGAGGGCC 640  
TAGATTTTTCTTCAGGAGGTGAACGTCTATGGAGTCACTG 680  
GCCAGGGCATGAAGGCAGGGCTGGGAATGGCAGCCCTAGTT 720  
CTGCGTCCCCCCCCACGCTTTGGACCTTATGCAGCTCTACA 760  
CCCACGTGTCTGAGAACTTTGCCACCTTATGCCCGGGCCCC 800  
810 820 830 840  
ATTCTCAGGCTCCAGGAGTCTTTGGCCACCACAGAGACC 840  
TTCAAAACAGCAGAAAGTTCCGGATGGCAAAATGAGGGCTTC 880  
ACCCACAGCACCCTGTCTGACCCACTGTACGTTCTGGACCA 920  
GGCTGTAGGTGGCTACCTGCCCTCACAACCTGCCCGGTAC 960  
AGCGCCCTCCTGGCAGGAAACCTTCGAAATCTGAGAACTTC 1000  
1010 1020 1030 1040  
CACACCTGAGGCACCTGAGAGAGGAACTCTGTGGGGTGGG 1040  
GGCCGTTGCAGGTGTACTGGGCTGTGAGGGATCTTTTCTA 1080  
TACCAGAACTGCGGTCACTATTTTGTAAATAAATGTGGCTG 1120  
GAGCTGATCCAGCTGTCTCTGACAAAAAATAAAAAAAAAA 1160  
AAAGGGGGGGCCGC 1173

Fig. 48

10	20	30	40
KFSAGGFWECCQHRVTVFQYIGELCRYLVNCFPSKAERG 40			
HKVRLAVGSGLRFPOTWERFVRRFGFLQVLESTYGLTEGNVA 80			
TINYTGQRGAVGRASWLYKHIFPFSLRVQVTTGEPIROP 120			
QGHCMATSPGEPGLLVAFVSQGSFFLGYAGGPFLAQGKLL 160			
KDYFRFGDYFFNTGOLLVCCGCGFLRFHRTGDTFRWKGE 200			
210	220	230	240
NVATTEVAEVFEALDFLCEVNVYGVTVPGHEGRAGMAALV 240			
LRFFHALGLMGLYTHVSENLPFYARFRFLRLGESLATTET 280			
FKOCKVRMANEGDFSTLSOPLYVLDCAVGAYLPLTTARY 320			
SALLAGNLR 330			

Fig. 49

10 20 30 40  
CGACCCACGGCTCCGGGCGGGCGGGGCGGGCGGGCGGGCG 40  
GGGCTGGCGGGGCGGGCGGGCCATGCAGGGCGCAGAGCCG 80  
GCTAAACCCCTGCTGAGACCCGGCTCCGTGCGTCCAGGGGC 120  
GGCTAATGCCCTCAGGCTGTCTACGCTGCTGCAACCGGG 160  
CCGCATCTGGACGGGGCGCCGCGCGGGCGGAGCCGACGCCG 200  
210 220 230 240  
GGCCACAATGCTGCTTGGAGCCTCTCTGGTGGGGGTGCTG 240  
CTGTTCTCCAAGCTGGTGTGAAACTGCCCTGGACCCAGG 280  
TGGGATTCTCCCTGTGTCTCTACTTGGGATCTGGCGG 320  
CTGGCGCTTCATCCGGGTCTTCATCAAGACCATCAGGCGC 360  
GATATCTTTGGCGGCTGGTCTCTCTGAAGGTGAAGGCAA 400  
410 420 430 440  
AGGTGGGACAGTGCCCTGCAGGAGCGGGCGGACAGTGGCCAT 440  
TTTGTCTTGCCTCTACCGTTCGGGCGCCACCCCGACAAGACG 480  
GCCCTGATCTTCGAGGGCACAGATAACCCACTGGACCTTCC 520  
GCCAGCTGGATGAGTACTCAAGCAGTGTAGCCCACTTCCT 560  
GCAGGCCCCGGGGCCTGGCCTCGGGCGGATGTGGCTGCCATC 600  
610 620 630 640  
TTCATGGAGAACCAGCAATGAGTTGCTGGGGCTATGGCTGG 640  
GCATGGCCCAAGCTCGGTGTGGAGGGCAGCCCTCATCAACAC 680  
CAACCTGCGGGCGGGATGCTCTGCTGCACTGGCTCAACACC 720  
TGGCGCGCACGGGGCCTTGTCTTTGGCAGCGAAATGGCCT 760  
CAGCCATCTGTGAGGTCCATGCCAGCCTGGACCCCTCGCT 800  
810 820 830 840  
CAGCCTCTTCTGCTCTGGCTCTCTGGGAGCCCGGTGGGGTG 840  
CCTCCAAGCACAGAACACCTGGACCCCTCTGCTGAAAGATG 880  
CTCCCAAGCACCTTCCCAGTTGCCCTGACAAAGGGCTTCAC 920  
AGATAAATCTGTTCTACATCTACACATCCGGGCACCAAGGG 960  
CTGCCCCAAGGCGGCCATCGTGGTGCACAGCAGGTATTACC 1000  
1010 1020 1030 1040  
GCATGGCTGCCCTGGTGTACTATGGATTCCGCATGCGGCC 1040  
CAACGACATCGTCTATGACTGCCCTCCCCCTCTACCACTCA 1080  
GCAGGAAACATCGTGGGAATCGGCCAGTGCCTGCTGCATG 1120  
GCATGACGGTGGTGAATCGGAAGAAAGTTCTCAGCCTCCCG 1160  
GTTCTGGGACGATTGTATCAAGTACAACAGCAGGATTGTG 1200

Fig. 50A

1210	1220	1230	1240
CAGTACATTGGTGAAC	TGTGCCGCTACCTC	CTGAACCAGC	1240
CACCGCGGGAGGCAG	AAAAACCAGCACC	AGGTTCCGATGGC	1280
ACTAGGCAATGGCTCC	GGCAGTCCATCTG	SACCAACTTT	1320
TCCACCGCTTCCACAT	ACCCCAGGTGGCT	GAGTTCTACG	1360
GGGCCACAGAGTGCA	AACTGTAGCCTGG	GCAACTTCGACAG	1400
1410	1420	1430	1440
CCAGGTGGGGGCTGT	GGTTTCAATAGCC	GCATCCTGTCC	1440
TTGGTGTACCCCATC	CGGTTGGTACGT	GTCAACGAGGACA	1480
CCATGGAGCTGATCC	GGGGGGCCGACGG	CGTCTGCATTCC	1520
CTGCCAGCCAGGTG	AGCCGGCCAGCTG	GGTGGGCCGCATC	1560
ATCCAGAAAGACCCC	CTGGCCGCTTCG	ATGCTACCTCA	1600
1610	1620	1630	1640
ACCAGGGGGCCAAACA	ACAAGAAGATTGCC	AAGGATGTCTT	1640
CAAGAAGGGGGACCA	GGCCTACCTTACT	GGTGATGTGCTG	1680
GTGATGGACGAGCTG	GGGCTACCTGTAC	TTCGAGACCGCA	1720
CTGGGGACACGTTCC	GCTGGAAAGGTGA	GAACGTGTCCAC	1760
CACCGAGGTGGAAGG	CACACTCAGCCGC	CTGCTGGACATG	1800
1810	1820	1830	1840
GCTGACGTGGCCGTG	TATGGTGTGAGGT	GCCAGGAACCG	1840
AGGGGCGGGCCGGA	ATGGCTGCTGTGG	CCAGCCCCACTGG	1880
CAACTGTGACCTGG	AGCGCTTTGCTCA	SGTCTTGGAGAAG	1920
GAAGTGGCCCTGTAT	GGCGGCCCCCATC	TTCTTGGCCCTCC	1960
TGCCTGAGCTGCACA	AAAACAGGAACCT	ACAAGTTCCAGAA	2000
2010	2020	2030	2040
GACAGAGCTACGGA	AGGAAGGGCTTTG	ACCGGGCTATTG	2040
AAAGACCCGCTGTTC	TATCTAGATGCC	CAGAAAGGGCCGT	2080
ACGTCCCGCTGGAC	CAAGAGGGCTTAC	AGCCGCATCCAGGC	2120
AGGCGAGGAGAAGCT	GTGATTCCCCCAT	GGCTCTGAGGG	2160
CCGGCGGATGCTGG	ATCCGAGCCCCAG	GTTCGCCCCAG	2200
2210	2220	2230	2240
AGCGGTCTCTGGACA	AGGGCCAGACCA	AAAGCAAGCAGGG	2240
GGCAGCTCCATGCTG	AGGTGCTGGCCCT	CCATCCAAAAC	2280
GGCAAGTGACTCAT	TGGCTTCCCAAC	CCCTTCCAGAGG	2320
TCTGTGAAAGTCTC	ATGTGTCCAAGT	TCCGTCTCTGGG	2360
GGCAGGCCCTCTG	GTTCGCCAGGCT	GAAGTGAAGGTT	2400
2410	2420	2430	2440
CTCAGGATGATGTCT	TGGGTGAGGGT	AGGGAGAGGACA	2440
GGGTACCCGAGCCCT	TCCCAGAGAGC	AGGGAGCTTATA	2480
TGGAACCAAGAGC	AGAAGTCCCCAG	ACTCAGGAAGTCA	2520
GAGTGGGCAGGGAC	AGTGGTAGCATCC	ATCTGGTGGCCAA	2560
AGAGAAATGCTAG	CCCCAGAGCTG	CCCCAAGTTCACT	2600

Fig. 50B

2610	2620	2630	2640
.....	.....	.....	.....
CCACCCCCACCTCCAGGAGGGGAGGAGAGGACCTGACATC	2640		
TGTAGGTGGCCCCCTGATGCCCCATCTACAGCAGGAGGTCA	2680		
GGACCACGCCCCCTGGCCTCTCCCCACTCCCCCATCCTCCT	2720		
CCCTGGGTGGCTGCCTGATTATCCCTCAGGCAGGGCCTCT	2760		
CAGTCCTTGTEGGTCTGTGTCACCTCCATCTCAGTCTTGG	2800		
2610	2820	2830	2840
.....	.....	.....	.....
CCTGGCTATGAGGGGAGGAGGAATGGGAGAGGGGGCTCAG	2840		
GGGCCAATAAACTCTGCCTTGAGTCCTCCTAAAAAAAAAA	2880		
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	2907		

Fig. 50C

10	20	30	40
.....	.....	.....	.....
MLLGASLVGVLLFSKLVKLFWTQVGFSLFLYLGSGGWR	40		
FIRVFIKTIIRQIFGGLVLLKYKAKVRGCLCERRTVPIIF	80		
ASTVRRHPDKTALIFEGTDTHWTFRQLDEYSSSVANFLOA	120		
RGLASGGVAAIFMENRNEFVGLWLGMAKLGVEAALINTNL	160		
RRQALLHCLTTSRARALVFGSEMASAICEVHASLOPSSL	200		
210	220	230	240
.....	.....	.....	.....
FCGGSWEFGAVFPSTEHLDPLLKDAPKHLFSCPDKGFTCK	240		
LFYIYTEGTTGLPKAAIVVHSRYRMAALVYYGFRMRPNC	280		
IYYDCLPLYHSAGNIVGIGQCLLHGMTVVIRKKFSASRFW	320		
DDCIKYNCTIVQYIGELCRYLLNGFPFRAENQHGVMALE	360		
NGLRGSIWTFNSSRFHIFGVAEFYGATECNCSLGNFDSGV	400		
410	420	430	440
.....	.....	.....	.....
GAQGFNSRILSFVYPIRLVVRVNEDTMELIRGFDGVCIFCG	440		
PGFPGOLVGRITCKQPLRRFDGYLNGGANNKKIAKQVFXK	480		
GCGAYLTGCVLVMDLGYLYFRORTGOTFRWKGENYSTTE	520		
VEGTLERLLDMADVAVYGVEVPQTEGRAGMAAVASPTGNC	560		
DLRFACVLEKELPLYARPIFLRLLPFLHKTGTXYKPKTE	600		
610	620	630	640
.....	.....	.....	.....
LRKEGFQPAIVKIFLFYLDACKGRYVFLOCEAYSRICAGE	640		
EKL	640		

Fig. 51

GTCGTTGGGATCCTCGGCTGCTTAGATCTCGGAGCCACCTGTGTTCTGGCCCCCAAG  
TTCTCTACTTCCTGCTTCTGGGA  
TGACTGTCCGCAGCATGGCGTGACAGTGATCCTGTATGTGGGCGAGCTCCTGCGATA  
CTTGTGTAACATTCCCCAGCAAC  
CAGAGGACCGGACACATACAGTCCGCTGGCAATGGCAATGGACTACGGGCTGAT  
GTGTGGGGAGACCTTCCAGCAGCG  
TTTCGGTCTTATTTCGGATCTNCGGAAGTCTTACGGGCTTCCACAGAAGGGCAACAT  
GGGGCTTTAGTTCAAATATTGTT  
GGGGGCGCTGCGGGGCGCTGGGGGCAAAGATGGAGCTTGCTCCTCCGAATGCTGT  
CCCCCTTTGAGCTGGTGCAGTTCC  
ACATGGAGCGCGCGGAGCCTGTGAGGGACAATCAGGGCTTCTGCATCCCTGTAGGG  
CTAGGGGAGCCGGGGCTGCTGTTG  
ACCAAGGTGGTAAGCCAGCAACCCTTCGTGGGCTACCGCGGCCCCCGAGAGCTGTC  
GGAACGGAACCTGGTGGCAACGT  
GCGGCAATCGGGCGACGTTTACTACAACACCGGGGACGTACTGGCCATGGACCGCG  
AAGGCTTCCTCTACTTCCCGGACC  
GACTCGCGGACACCTTCCGATGGAAGGGCGAGAACGTGTCCACGCACGAGGTGGAG  
GCGGTGTTGTGCGACGGTGGACTTC  
TTGCAACAGGTTAACGTGTATGGCGTGTGCGTGCCAGGTTGTGAGGGTAAGGTGGC  
ATGGCTGCTGTGGCATTAGCCCC  
CGGCCAGACTTTTCGACGGGGAGAAAGTTGTACCAGCACGTTCCGGCTTGGCTCCCTGC  
CTACGCTACCCCCCATTTATCC  
GCATCCAGGACGCCATGGAGGTACCCAGCACGTTCAAACTGATGAAGACCCGGTTG  
GTGCGTGACGGCTTCAATGTGGGG  
ATCGTGCTTGACCCCTCTGTTTGTACTGGACAACCGGGGCCAGTCCTTCCGGCCCCCTG  
ACGGCAGAAATGTACCAGGCTGT  
GTGTGAGGGAACCTCGAGGCTCTGATCACCTGGCCAACCCACTGGGGTAGGGATCA  
AAGUCAGCCACCCCCACCCCAACA  
CACTCGGTGTCCCTTTTCATCCTGGGCTGTGTGAATCCAGCCTGGCCATACCTCA  
ACCTCAGTGGGCTGGAAATGACA  
GTGGGCCCTGTAGCAGTGGCAGAAATAAACTCAGMTGYGTTACAGAAA

Fig. 52

10	20	30	40
VVGILGCLDLGATCVLAFKFSTSCFWDDCRQHGVTVILYV	40		
GELLRYLCNIPGGFEDRTHTVRLAMGNGLRADVWGDLFAA	80		
FRSYFGSXEVLRASTECCGHALVGILLGALRGFGGKDGAC	120		
LLRMLSPFELVCFDMEAAEPVRONGGFCIPVGLGEFGLLL	160		
TKVVSOGPFVGYRGPRELSEKLVNVRQSGGVYYNTGDV	200		
210	220	230	240
LAMCREGFLYFRDRLGOTFRWKGENVSTHEVEGVLSGVDF	240		
LGGVNVYGVGVFGCEGKVGMAAVALAPGGTFDGEKLYCHV	280		
RAWLPAYATPHFIRIQDAMEVTSTFKLMKTRLVREGFNVG	320		
IVVDPLFVLONRAQSFRLTAEMYQAVCEGTWRL	360		

Fig. 53



10 20 30 40  
AACGGCAAGTAAGCGCAACGCAATTAATGTGAGTAGCTCA 40  
CTCATTAGGCACCCCGAGGCTTTACACTTTATGCTTCCGGG 80  
CTCGTATGTTGTGTGGAATTGTGAGCGGATACCAATTTCA 120  
CACAGGAACCAGCTATGACATGATTACGAATTTAATACGA 160  
CTCACTATAGGGAATTTGGCCCTCGAGGCCAAGAATTCGG 200  
210 220 230 240  
CACGAGGGGTGCTGAGCCCCGTGCGCGGTTTCTGGTGCCTA 240  
GAGACTGTAAATCGCTGCGCTTCTCAGTCATCATCATCCC 280  
AGCTTTTCCCGGCTCGAATTCAGCCTCCAAGCTCAAGCTCG 320  
CGGGAAGAGACTACCTGAGAGGAGAAAAGCTTCTGTCCCTG 360  
GACCTTCTTCTGAGGGTGGAGTGGGAGGCTCCCTGCTTTG 400  
410 420 430 440  
CAGCCGCCCCAGTGACCCCAAGCTTAACTCTTCAGCACCACTT 440  
GGGGCGACCTTTTCGGTGCAGAACCTACGATTCTGTTTCTC 480  
AGGATTCTTCCCATCCCGCTTCGCCCGCGAAGGCTGAC 520  
AAGAACTTCAGGTGTAAAGCCCTGAGTAGTGAAGATCTGGG 560  
GTCTCCGTGGAGAGCTGTGCCTGGAAGAGAAAGGAGCTGG 600  
610 620 630 640  
TGGGGGCTGAGATCAGAGCTGTCTTCTGGCCCAAGTTGCC 640  
CCATGCTTCTGTCTAGGCTAACAGTTCTAGGGGCTGGAA 680  
GGTCGTCTGCACTTCTTGCAGAAACTCTCTTCCCTTAC 720  
TTTTGGGATGACTTCTGCTTCTGTGTTGAAGGTGGTGTCA 760  
TTATAATTGGCTGAAGGAAGTATGAAAAAGAGAGGGAGCT 800  
810 820 830 840  
GGTGACTGTGCTGGATAAAATTCCTTGAGTCATGCCAAAAGA 840  
CAACCTCGGAAACCTTTTCATCATCTATGAGGGAGACATCT 880  
ACACCTATCAGGATGTAGACAAAAAGAGCAGCAGAGTGGC 920  
CCATGTCTTCTCTAACCATTCTCTCTCTGAATAAGGGGAGC 960  
ACGGTGGCTCTGCTGATGAAGCAATGAGCCGGACTTCGTTG 1000  
1010 1020 1030 1040  
ACGTGTGGTTCCGGCTCGCCAAAGCTGGGGCTGGGTGGTGGC 1040  
CTTTCTCAACACCAACATTTCGCTCCAACTCCCTCCTGAAT 1080  
TGCATCCCGGCTGTGGGCCCGAGAGCCCTAGTGGTGGGGCG 1120  
CAGATTTGCTTGGAAACGGTAGAAGAAATCCTTCCAAAGCCT 1160  
CTCAGAAAATATCAGTGTTTGGGGGATGAAGATTCTGTT 1200

Fig. 54A

1210	1220	1230	1240
CCACAAGGTGTAATTTCACTCAAAGAAAACTGAGCACCT	1240		
CACCTGATGAGCCCGTGCCACGCAGCCACCATGTTGTCTC	1250		
ACTCCTCAAGTCTACTTGTCTTTACATTTTTACCTCTGGA	1320		
ACAACAGGTCTACCAAAAAGCAGCTGTGATTAGTCAGCTGC	1360		
AGGTTTTAAGGGGTTCTGCTGTCCFGTGGGCTTTTGGTTG	1400		
1410	1420	1430	1440
TACTGCTCATGACATTGTTTATATAACCCCTTCTCTGTAT	1440		
CATAGTTCAGCAGCTATCCTGGGAATTTCTGATGTGTTG	1480		
AGTTGGGTGCCACTTGTGTGTTAAAGAAGAAATTTTCAGC	1520		
AAGCCAGTTTTTGGAGTGAAGTCAAGAAGTATGATGTGACT	1560		
GTGTTTCAGTATATTGGAGAAGCTTTGTGCTACCTTTGCA	1600		
1610	1620	1630	1640
AACAATCTAAGAGAGAAAGGAGAAAAGGATCATAAGGTGGG	1640		
TTTGGCAATTGGAAATGGCATACGGAGTGATGTATGGAGA	1680		
GAATTTTTAGACAGATTTGGAAATATAAAGGTGTGTGAAC	1720		
TTTATGACAGCTACCGAATCAAGCATATCTTTTCATGAAGTA	1760		
CACTGGGAGAAATGGAGCAATTTGGGAGAACAAATTTGTTT	1800		
1810	1820	1830	1840
TACAACTTCTTTTCCACTTTTGAAGTTAATAAAATATGACT	1840		
TTTCAGAAAGATGAACCCATGAGAAATGAGCAGGGTTGGTG	1880		
TATTCATGTGAAAAAAGGAGAAAGCTGGACTTCTCATTTCT	1920		
CGAGTGAATGCAAAAAATCCCTTCTTTGGCTATGCTGGGC	1960		
CTTATAAGCACACAAAAAGACAAATTTGCTTTGTGATGTTTT	2000		
2010	2020	2030	2040
TAAGAAAGGAGATGTTTACCTTAATACTGGAGACCTTAATA	2040		
GTCCAGGATCAGGACAAATTTCTTTTATTTTGGGACCGTA	2080		
CTGGAGACACTTTTCAGATGGAAAAGAGAAATGTGGCAAC	2120		
CAGTGAGGTTGCTGATGTTATTGGAAATGTTGGATTTTCTA	2160		
CAGGAAGCAAAACGTCTATGGGTGTGGCTATATCAGGTTATG	2200		
2210	2220	2230	2240
AAGGAAGAGCAGGAATGGCTTCTATTAATTTAAAAACCAA	2240		
TACATCTTTAGATTTGGAAAAAGTTTATGAACAAAGTTGTA	2280		
ACATTTCTACCAGCTTATGCTTGTCCACGATTTTAAAGAA	2320		
TTTACGAAAAAATGGGAAGCAACAGGAACATTCAAACTATT	2360		
GAAGCATCAGTTGGTGGGAAGATGGATTTAATCCACTGAAA	2400		
2410	2420	2430	2440
ATTTCTGAACCACTTTTACTTTCATGGATAACTTTGAAAAAGT	2440		
CTTATGTTCTACTGACCAGGGAAGCTTTTATGATCAAAATAAT	2480		
GTTAGGGGAAATAAACTTTAAGATTTTTATATCTAGAAC	2520		
TTTCATATGCTTTCTTAGGAAGAGTGAGAGGGGGGTATAT	2560		
GATTCCTTATGAAAATGGGGAAAGGGAGCTAACATTAATTA	2600		

Fig. 54B

---

2610	2620	2630	2640
.....			
<hr/>			
TGCATG	TAATTT	CCTTAAT	ATGAGAGATAATTTTTT 2640
AATTGC	AAGAATTT	TAATTT	CTTTTAATTGATATAAAC 2660
ATTAGT	TGATTATT	CTTTT	TATCTATTTGGAGATTCAGTG 2720
CATAACT	AAGTATTT	CCTTAATA	CTAAAGATTTTAAATA 2760
ATAA	ATAGT	GGCTAG	CGGTTTGGACAATCACTAAAAATGT 2800
.....			
2810	2820	2830	2840
.....			
<hr/>			
ACTTTCT	AATAAGT	AAAAATTT	CTAATTTTGAATAAAAGAT 2840
TAAATTT	TACTG	AAAAAAAA	AAAAAAAAAAAAAAAAAATTGGCG 2880
GCCGC	2885		

Fig. 54C

10	20	30	40
MLLSWLT	TVLGAGMV	VLHFLCK	LLFFYFW
40	DOFWFV	LKVVLI	
IIRLK	YEKRGEL	VTVL	DKFLSHAK
80	RQPRK	FFII	YEGGIY
TYG	VEKRSS	RYAHV	FLNHSS
120	SLKK	GOTV	ALLMS
VWF	GLAKL	GCVV	AFLENTN
160	IRS	NSSL	NCIRAC
DLL	GTVEE	ILP	SLSEN
200	ISV	WGMK	DSVP
210	220	230	240
FOEP	VFRSH	VVSL	LKSTCLY
240	IFT	SGTT	GLPKAA
VLR	GSAY	LWAF	GCTAHD
280	IVYI	TLFL	YHSSAA
LGAT	CVL	KKKFS	ASGFW
320	SOCK	KYDV	TVFQY
OSK	REGE	KDHK	VRLAIG
360	NGIR	SOVW	REFLDR
YAAT	ESSI	SFMN	YTRIGA
400	IGRT	NLFY	KLLSTF
410	420	430	440
CKCE	PMNE	GGWC	IVHKK
440	GE	PGLI	ISRV
YKHT	KOKL	LCOV	FKKG
480	SVYL	NTGOL	IVGDC
GOT	FRWK	GENV	ATTEVA
520	ADV	IGML	DFIG
GRAG	MAS	ILK	FN
560	SLD	LEK	VYEC
CEK	MEAT	GT	FKLL
600	KHQL	VED	GFN
610	620	630	640
YVLL	TRELY	CCIN	LGEIKL
619			

Fig. 55

10 20 30 40  
AAGTTCCCACTCCAGACTTCTGCGAGAACCCCGTGAGSAAG 40  
CAGCGAGAACCAGGGGTTTTCGAAGCCAGAGAAGGATGCGG 80  
ACTCCGGGAGCAGGAACAGCCTCTGTGGCCTCATTGGGGC 120  
TGCTTTGGCTTCTGGGACTTCCGTGGACCTGGAGCGCGGC 160  
GGCGGCGTTCGGTGTGTACGTGGGTAGCGGTGGCTGGCGA 200  
210 220 230 240  
TTTCTGGCTATCGTCTGCAAGACGGCGAGGGCAGACCTCT 240  
TTGGCCTCTCTGTCTGTGATCCGCGTGGGGCTAGAGCTAG 280  
ACGACACCGGGGAGCAGGAGACACGATCCACCGCATCTTC 320  
CAGGCGGTGGGCCAGCGACAGCCGGAGCGCCTGGCGCTGG 360  
TAGATGGGAGTAGCGGTATCTGCTGGACCTTCGCACAGCT 400  
410 420 430 440  
AGACACCTACTCCAATGCTGTGGCCAACTCTGTTCTTCCAG 440  
CTGGGCTTTGGGCCAGGCGATGTGGTGGCTGTGTTCTTGG 480  
AAGGCGGGCGGAGTTCTGTGGGACTGTGGCTGGGCTTGGC 520  
CAAGGCGGGTGTAGTGGCTGGGCTTCTCAATGTCAACCTG 560  
AGGCGGGGAGCCCCCTTGGCTTCTGCTTGGGCACATCAGCTG 600  
610 620 630 640  
CCAAGGCGCTCATTTATGGCGGGGAGATGGCAGCGGCGGT 640  
GGCGGAGGTGAGTGAGCAGCTGGGGAAAGAGCCTGCTCAAG 680  
TTCTGCTCTGGAGATCTGGGGGCTGAGAGCGTCTTGGCTG 720  
ACACGCGAGCTTCTGGACCCCATGCTTGTCTGAGGGCGCCAC 760  
CACACCCCTGGCACAGGCCCCAGGCAAGGGCATGGATGAT 800  
810 820 830 840  
CGGCTATTTTACATCTATACTTCTGGGACCACCGGACTTC 840  
CTAAGGCGGGCATTGTGGTGCACAGCAGGTACTACCGCAT 880  
CGCAGCCTTGGGCCACCATTCCTACAGCATGGCGGCGCAAC 920  
GATGTGCTCTATGACTGGCTAGCTCTCTACCACTCAGCAG 960  
GGAACATCATGGGCGTGGGACAGTGTATCATCTACGGGTT 1000  
1010 1020 1030 1040  
AACGGTGGTACTGCGCAAGAAGTTCTCGGCCAGCGGCTTC 1040  
TGGGACGACTGTGTCAAAATATAATTGCACGGTAGTGAGT 1080  
ACATCGGTGAAATAAGCCGCTACCTGCTAAGGCAGCCGGT 1120  
TCGCGATGTAGAGCGGGCGGCACCGCGTGGCGCTGGCGGTG 1160  
GGTAACGGACTGCGGCCAGCCATCTGGGAGGAGTTACAGC 1200

Fig. 56A

61/148

1210	1220	1230	1240
AGGGTTTCGGTGTGCGACAGATTGGCGAGTTCTACGGCGC	1240		
CACCGAATGCAACTGCAGCATTGCCAACATGGACGGCAAG	1280		
GTGGGCTCCTGCGGCTTCAACAGCCGTATCCTCACGCATG	1320		
TGTACCCCATCCGTCTGGTCAAGGTCAACGAGSACACGAT	1360		
GGAGCCACTGAGGGGACTCCCAAGGCCTCTGCATCCCGTGC	1400		

1410	1420	1430	1440
CAGCCCGGGGAACCTGGGGCTTCTCGTGGGGCCAGATCAACC	1440		
AGCAAGACCCCTCTGCGGCGCTTCGATGGCTATGTTAGTGA	1480		
CAGCGCCACCAACAAGAAGATTGCCACAGCGTGTTCCTGA	1520		
AAGGGGGACAGCGCCTACCTTTGAGGTGACGTGCTAGTGA	1560		
TGGACGAGCTGGGGTACATGTACTTCCGTGACCGCAGCGG	1600		

1610	1620	1630	1640
GGATACCTTCCGATGGCGCGGCGAGAACGTATCCACCACG	1640		
GAGGTGGAAGCCGTGCTGAGCCGCGCTGTTGGGGCCAGACGG	1680		
ACGTGGCTGTGTATGGAGTGGCTGTGCCAGGAGTGGAGGG	1720		
GAAAAGCGGCATGGCGGGCCATTGCGAGACCCCGACAACCAG	1760		
CTGGACCCCTAACTCAATGTACAGGAATTGCGAAGGTTTC	1800		

1810	1820	1830	1840
TTGCATCCTATGCCCGAGCCCATCTTTCCTGCGTCTTCTGCC	1840		
CCAAGTGGATACAACAGGGCACCTTCAAGATCCAGAAAGACC	1880		
CGACTACAGCGTGAAGGGCTTTGACCCCGCGCCAGCCTCAG	1920		
ACCGGCTCTTCTTTCTAGACCTGAAACAGGGACGCTACCT	1960		
ACCCCTGGATGAGAGAGTCCATGCCCGCATCTGCGCAGGGC	2000		

2010	2020	2030	2040
GACTTCTCAGTCTGAGGCTGGTGAATGGGATGGCCCTGGA	2040		
CTTGTGAGAGCAGGGAGCCGGACACCCCTGTTCAAGGTT	2080		
TCTCCTGCTGGCCAGGTGGCCAGCAGCACTGTGGGTGC	2120		
AGGAAGCTGGAACCTGAGTGGCGGGGTGTCCCTTTCTTAC	2160		
AACCCACCATGCACACATCTAGCCTCTGCTTGGTCTTTT	2200		

2210	2220	2230	2240
TCTCCATCTCTTTCTCCGTGCCCGAGCAGGAGCCCGACAG	2240		
ACACATTGGCTGCTGTGTCTCTGCAGTGGGACCGGTGTCTA	2280		
GGGGTCCATGCTGCAAGCTGTGACCGGCACTGGTGGCCAC	2320		
CTCCCTTCCCGATTGTGCTTTAGGTTCCTCCACTGTGCGC	2360		
CGGTGAAGCAAGTGGGGACCCACATAGCTGTGTGCTCTGC	2400		

2410	2420	2430	2440
TGAGGGTTGGTAGCAAAATGCACCCCTCATGTGAGCTGGGAG	2440		
ACACATGCAGTCTCCCACTGACCCCGCAATCAACTGAAGAT	2480		
ACTGTTTTGTATTATTGTTTTGAGATAGGGTCTCACTGTG	2520		
GAGGCCAAGCTGGCCTCAGGCTCACCCTCTACTGCTCTC	2560		
GGGCACCGAGCTGCAGTTTGATGACATGTATGCACTATTG	2600		

Fig. 54B

```
      2510      2620      2630      2640
.....|.....|.....|.....|
TTCTAAGGGTCTTCTGAGTCCCTGCTTTCCCTCATGTCC 2640
TAAACCTTCCAGAACTGACTCTGATCACTTGGATGTAGC 2680
TAGTGTTGGCCCTGCCCACGTGTGTCAATTCAGGGGTCCC 2720
CAGGCATCATCTCTGGAGGCCCTAACCTTGGCAAAGCTTG 2760
GATGTCTCACATCACAGCAGGAGACCCAGGAAGGTTGCT 2800

      2810      2820      2830      2840
.....|.....|.....|.....|
GTGGTGTCTCTTGGGCACCCCTGGCGGCAGCCGTGGACAT 2840
GCTTCCCTGCTGTGATAGCCCAAACCTGTTGCCTATGACAT 2880
TTGAGGTCTACCCTTCTGGCTGCCATGGTCCCCATTGAGA 2920
TCTTTGGTGACTCACCTCAGCCACCAAGCCAGGCCTCTGC 2960
CTTCCTTCAGCTCTAAGGGGATGAAGGGTGTGGACAGAGC 3000

      3010      3020      3030      3040
.....|.....|.....|.....|
AGCCACAGGCTGCCCCACAGTCACCCACATGCAAGTETTAT 3040
TTCCTTGTTTGTGTTTTAAAAAAATAAACATGCTGAGCCTTG 3080
AAAAAAAAAAAAAAAAAAAA 3096
```

Fig. 56C

10	20	30	40
MRTPGAGTASVASLGLLWLLGLPWTWSAAAAFGVYVGS	GG	40	
WRFLRIVCKTARROLFGLSVLIRVRLRLRRHRRAGSTIPR	80		
IFQAVACRQPERLALVDASSGICWTFAGLOTYSNAVANLF	120		
LQLGFAPGGVYAVFLEGRPEFVGLWLGLAKAGVVAALLNV	160		
NLRREPLAFCLGTSAAKALIYGGEMAAAYAEVSECLGKSL	200		
210	220	230	240
LKFCSGDLGPESVLPOTCLLOFMLAEAPTTFACAPGKGM	240		
DGRIFYIYTSGETTGLPKAAIVVHSRYRRAAFGHHSYGM	280		
ANDVLYGCLFLYHSAGNIMGVGGCIYGLTVVLRKKFSAS	320		
RFWDGCVKYNCTVYGYIGEICRYLLRQFVRQVERRHRVRL	360		
AVGNGLRPAIWEFTGGFGVRQIGEFYGATECNCSTANMC	400		
410	420	430	440
GKVGSGGFSNRILTHVYFIRLVKYNEOTMEPLRCSGGLC	440		
PCGPGEFGLLVGGINCCDPLRRFDGYVSDSATNKKIAHSV	480		
FRXGDSAYLSGGVLYMDELGYMYFRORSGDTFRWRGENVS	520		
TTEVEAVLSRLLGOTQVAVYGVAVPGVEGKSGMAAIAQPH	560		
NGLDFNSMYCELCKVLAAYACFIFLRLLPCVDTTGTGFKI	600		
610	620	630	640
KTRLQREGFDFRQTSORLFFLDLKGGRYLFLDERVHARIC	640		
AGGFSL	646		

Fig. 57



10 20 30 40  
GACACAGTACTGCCGATGTTGGACAGAGGATCGCTTAACA 40  
GAACGAAATCTCAAAACAAATTAACAGGACCGGTTGCTT 80  
GATTTCCCAAAATCAGAAAAGGCTCGAAATGTCTAGAGGGG 120  
CTGACTGATGCAGCGGTGACCCGGACTGGAGACAGTTGGA 160  
CGCGATCATCTCTGGTGCTTTTGTTC AACCTTGAAACCTT 200  
210 220 230 240  
CGCCACAGGAGACTTGCCTGAGCAGAGAAAGCAAACGTGGA 240  
GAAACAAAAGAGAGATCTAGCGAAAAGCCTCTGGGACCAAG 280  
GAGGGGAGGTGGGACTCTGGGTTGGCGGTGGACCTGCTG 320  
CCGGCTATTAAATAATAGGGTCCGGATGCGTTTATAAGGTG 360  
TTTGATTAAACAAAAGACTCTATGAGAGAAAGAACTAGC 400  
410 420 430 440  
AACAGCCCCACGTCTGAGTCGTGCGCTCCGACCTTTTTCA 440  
ACGTGGGTTCTTTGGGCGGAGCGTCTGTTTGGGAGAACTA 480  
GATCTCACCTGACCCCAAGCTGAAAACAAAGCGCTGTGG 520  
CATCTGGGGCCACCCAAAGCTGACAAGGGCGCGCCCTCTGA 560  
GCACACGAGGTGCCCCACGAGGGGGAGGGACCCACAGCCG 600  
610 620 630 640  
TCCCGCCCGCACCGGGGTGTCCGCTGCGGGGACCTGCAAC 640  
CGAGTCGCGCACCCCTAGTCGCGAGCGCGTCCCGCGCGGAA 680  
CCCGGTCGTCAGCTCGTCAGCACTGCTCTCTCTCTCTCT 720  
CGCCCGCGCGCGCGCTGCAAGGCTCGAGCGCTTCTCTCGGC 760  
CCCGCGCGGGGACCGGGGACCCCGCAGCCATCGCCATGCTG 800  
810 820 830 840  
CCTGTGCTCTACACCGGCTGGCGGGGCTGCTGCTGCTGC 840  
CTCTGCTGCTCACCTGCTGCTGCCCCCTACCTGCTCCAGGA 880  
CGTGGGTTCTTCTTGCAACTGGCCAAACATGCCCCGGCAG 920  
GTGCGCAGCTACCGGCAGCGGGGACCCGTGCGGACCATCC 960  
TGCATGTCTCTTGGAGCAAGCGCGCAAGACCCCGCACAA 1000  
1010 1020 1030 1040  
GCCCTTCCTGCTGTTTTGCGGACGAGACGCTTACCTACGCC 1040  
CAGGTAGACCGGCGCAGCAACCAAGTAGCGCGAGCGCTGC 1080  
ATGATCACCTGGGCCTGCGGCAGGGGGATTGCGTGGCCCT 1120  
CTTCATGGGCAATGAGCCGGGCTACGTGTGGCTCTGGCTG 1160  
GGACTGCTCAAACTGGGGCTGTCCCATGGCGTGGCTCAACT 1200

Fig. 58A

1210 1220 1230 1240  
ACAACATCCGTGCCAAGTCTCTGCTACACTGCTTTCAGTG 1240  
CTGCGGGGGCGAAGGTGCTGCTGGCCCTCCCCAGAGCTACAC 1280  
GAAGCTGTGAGGAGGTTCTTCCAACCGTGAAAAAGGAGG 1320  
GCCTGTCCGTCTTCTACGTAAGCAGAACTTCTAACACTAA 1360  
TGGCCTGGACACAGTACTGGACAAAGTAGACGGGGTGTCTG 1400

1410 1420 1430 1440  
GCGGACCCCATCCCGAGTCTGTGGAGGTCTGAAGTCACGT 1440  
TCACCACACCCGCGAGTCTACATATATACTTCGGGCACCAC 1480  
AGGTCTTCCAAAGGCTGCAACCATTAAATCACCATCGCCTC 1520  
TGGTATGGGACCAGCCTTGGCCTGAGGTCCGGAATTAAGG 1560  
CTCATGACGTCTCTACACCACCATGCCCCCTGTACCACAG 1600

1610 1620 1630 1640  
CGCGGCGCTCATGATTGGCCCTCCACGGATGCATTGTGGTT 1640  
GGGGCTACATTTGCTTTGCGGAGCAAAATTTTCAGCCAGCC 1680  
AGTTTGGGACGACTGCAGGAAATACAACGCCACTGTCTAT 1720  
TCAGTACATCGGTGAATCTGCTTCGGTACCTCTGCAACACG 1760  
CCCCAGAAACCAAAATGACCGGGACCACAAAGTGAATAATG 1800

1810 1820 1830 1840  
CACTAGGAAATGGCTTACGAGGAGATGTGTGGAGAGAGTT 1840  
CATCAAGAGATTTGGGGACATTACATTTATGAGTTCTAC 1880  
ECTTCCACTGAAGGCCAACATTGGATTTATGAAGTATCCAA 1920  
GAAAAATCGGAGCTGTGTGGAGAGAGAAAAATACCTACAAA 1960  
AAAAAGTTGTAAGGCACGAGCTGATCAAGTATGACGTGGAG 2000

2010 2020 2030 2040  
AAGGATGAGCCTGTCCTGTCATGCAAAATGATATTTGCATCA 2040  
AAGTCCCCAAGGAGAGGTTGGACTCTTTGATTTGCAAAAT 2080  
CACAGAGCTCAGACCATTTTCTTGGCTATGCTGGAGGAAAG 2120  
ACCCAGACAGAGAAAGAAAAAGCTCAGAGATGTTTTTAAGA 2160  
AAGGAGACGCTCTACTTCAACAGTGGCGATCTCTGTATGAT 2200

2210 2220 2230 2240  
CGACCGTGAAAAATTTTCATCTATTTTCACGACAGAGTTGG 2240  
GACACCTTCCGGTGGAAAGGAGAGAAATGTAGCTACCACGG 2280  
AAGTCCCTGACATTGTGGGACTGGTAGATTTTGTGGAAGA 2320  
AGTGAATGTTTACGGTGTGCCCCGTGCCAGGTCTATGAAGGT 2360  
CGCATCGGGATGGCTCTGATCAAGATGAAAAGAAAATACG 2400

2410 2420 2430 2440  
AGTTCAATGGAAGAACTCTTTTCAGCAGATCTCGGAGTA 2440  
CCTGCCCAGTTACTCGAGGCCCTCGGTTCTGAGAAATACAA 2480  
GATACCATTGAGATCACCGGGACTTTTAAACACCCGCAAG 2520  
TGACCCCTGATGGAAAGGGCTTTAAACCCCTCAGTCATCAA 2560  
AGATACCTTGTATTTTCATGGATGACACAGAAAAAACATAC 2600

Fig. 58B

---

2610	2620	2630	2640
.....			
<hr/>			
GTGCCC	ATGACTGAGGACATTTATAATGCCATAATTGATA	2640	
AGACTC	TGAAGCTCTGAATGTTGCCTGGCTCCTAACACTT	2680	
CCAGAA	AGAAACACAATAGGCCTAGCATAGCCCCTTCACA	2720	
TGTGTA	ATCCAACCTTTAACTTGATTAAAGGTTATAGGTGT	2760	
GATTTT	CCTAGGAAATTATTCATTTAAAGGACAATTGTT	2800	
2810	2820	2830	2840
.....			
<hr/>			
TGTTTG	TTTGTTTGTATTAATTACACCAGAACGTT	2840	
TGCAAG	TAAAAAGATTTAAAGTCACITATTTTCAATGTG	2880	
CACCTG	CCATTTGTCCTTGCAAACCTTAGCTTCTTGGAGAG	2920	
AGGGCC	TTATTTTTTTAAAGACATAATAAACTATGTAAAC	2960	
ACT	2963		

Fig. 58C

10	20	30	40
MLPVLVTGLAGLLLLPLLLTCCCPYLLQCVRFLLQLANMA	40		
RCVRSYRGRFPVRTILHYFLEQARKTPHKPFLFRDETTLT	80		
YACVDRRSNCGVARALHDLGLRGGOCVALFMGNFPAYVWL	120		
WLGLLKLGCPMACLNYNIRAKSLLHCFQCCGAKVLLASPE	160		
LHEAVEEVLPFTLKKEGVSIFYYSRTSNTNGVDIVLCKVGG	200		
210	220	230	240
VSAGPFPESWRSEVTFITPAVYIYTSGETTGLPKAATINHH	240		
RLWYGTSLALRSGIKAHQVIYTTMFLYHSAALMIGLHGCI	280		
VYSGATFALRSKFSASCFWDCCKYNATVICYIGELLRYLC	320		
NTPGKPNQROHKVIALGNGLRGQVWREFIKRFGGIHIYE	360		
FYASTEENIGFMNYPRKIGAVGRENVLCKKVVRHELKYG	400		
410	420	430	440
VEKDEPVRDANGYCIKVPKGEVGLLTCKITELTPFFGYAG	440		
GKTGTETKKKLRCVFKKGGVYFNSGOLLMDRENFIYFHR	480		
YSGTFRWKGENVATTEVADIVGLVDFVEEVNYYGVFVPGH	520		
SGRIGMASIKMKENYEPNGKKLFQHISEYLPSSYRRPFLR	560		
IGOTIEITGTFKRRKVTLMEEGPNFPIKOTLYPMOOTEX	600		
610	620	630	640
TYVPMTEDIYNATICKTKL	620		

Fig. 59

10 20 30 40  
GATCAGCTCTTCTATATCTACACGTCGGGCACCACGGGGC 40  
TACCCAAAGCTGCCATTGTGGTGCACAGCAGGTATTACCG 80  
AATGGCTGCCCTGGTGTACTATGGATTCCGCATCGGGCCT 120  
GATGACATTGTCTATGACTGCCCTCCCCCTCTACCACTCAG 160  
CAGGAAACATTGTGGGGATTGGCCAGTGCCTACTCCACGG 200  
210 220 230 240  
CATGACTGTGGTGATCCGGAAGAAGTTTTTCAGCCTCCCGG 240  
TTCTGGGATGACTGTATCAAGTACAACAGCACAATTGTAC 280  
AGTACATTGGTGAGCTTTGCCGCTACCTCCTGAACCAAGCC 320  
ACCCCGTGAGGCTGAGTCTCGGCACAAGGTGCGCATGGCA 360  
CTGGGCAACGGTCTCCGGCAGTCCATCTGGACCGACTTCT 400  
410 420 430 440  
CCAGCCGTTTCCACATTCCCAAGGTGGCCGAGTTCTACGG 440  
GGCCACCGAGTGCAACTGTAGCTTGGGCAACTTTGACAGC 480  
CAGGTGGGGGGCCTG7GGCTTCAATAGCCGCATCCTGTCT 520  
TTGTGTACCCCATCCGCTTGGTAAGAGTCAATGAGGATAC 560  
CATGGAACATGATCCGGGGACCCGATGGCGTCTGCATTCCC 600  
610 620 630 640  
TGTCAACCAGGCCAGCCAGGCCAGCTGGTGGGTGGCATCA 640  
TCCAGCAGGACCCCCCTACGCCGTTTTTGATGGCTACCTCAA 680  
CCAGGCTGCCAACAACAAGGATTGCTAGTGATGTCTTC 720  
AAGAAAGGGGACCAAGCCTACCTCACTGGTGAAGTGCTGG 760  
TGATGGATGAGCTGGGCTACCTGTACTTCCGAGAACCGCA 800  
810 820 830 840  
AGGGGACACGTTCCGCTGGAAAGGGGAGAAATGTGTCTACC 840  
ACTGAAGTGGAGGGGACACTCAGCCGCTGTCTCAGATGG 880  
CAGATGTGGCTGTTTTATGGTGTGAGGTGCCAGGAGCTGA 920  
GGGCCGAGCAGGAATGGCTGCTGTGGCAAGCCCCACTAGC 960  
AATGTGTACCTGGAGAGCTTTTGCACAGACCTTGAAGAAAG 1000  
1010 1020 1030 1040  
AGCTGCCCTGTACGCCCGCCCCATCTTCTCTCGCTTCTT 1040  
GCCAGAGCTGCACAAAACAGGAACCTTCAAGTTCCAGAA 1080  
ACAGAGTTGCCGAAGGAGGGCTTTGACCCGTCTGTGTGTGA 1120  
AAGACCCACTCTTCTATTTGGATGCCCGGACAGGCTGCTA 1160  
TGTTGCCTGGACCAAGAGGGCTATACCCGCATCCAGGCA 1200

Fig. 60A

---

1210 1220 1230 1240  
.....  
GGCGAGGAGAAGCTGTGATTTCCCCCACATCCCTCTGAGG 1240  
GCCAGAGGATGCTEGATTCAGAGCCCCAGCTTCCACTCCA 1280  
GAAGGGGTCTGGGCAAGGCCAGACCAAAGCTAGCAGGGCC 1320  
GECACCTTCACCCTAGGTGCTGATCCCCCT 1350

---

Fig. 60B

```
      10      20      30      40
.....
DCLFYIYTSGETTGLPKAAIVVHSRYRMAALVYYGFRMRP 40
DDIVYDCLFLYHSAGNIVGIGCCVLHGMTVVIRKKFSASR 80
FWDDCIKYNCTIVQYIGELCRYLLNQFPPEAESRHKVRMA 120
LGNGLRQSIWTOFSSRFHIPKVAEFYGATECNCSLGNFDS 160
QVGACGFNSRILSFVYFIRLVRVNEDTMELIRGPQGVCI 200
      210      220      230      240
.....
CGFGQPGQLVGRIIQQDPLRRFDGYLNCGANNKKIASDVF 240
KKGGQAYLTGDVLVMDLGYLYFRDRTGDTFRWKGENVST 280
TEVEGTLRLLQMAOVAVYGVVPGAEGRAGMAAVASPTS 320
NCOLESFAGTLKKELPLYARPIFLRFLPELHKTGTGTFCK 360
TELRKEGFQPSVVKQPLFYLDARTGCYVALQGEAYTRICA 400
      410      420      430      440
.....
GEEKL 405
```

Fig. 41

10 20 30 40  
ATGCGGGCTCCTGGAGCAGGAACAGCCTCTGTGGCCTCAC 40  
TGGCGCTGCTTTTGGTTTCTGGGACTTCCGTGGACCTGGAG 80  
CGCGCGCGCGCGCGTTCGTGTGTACGTGGGTGGCGCGCGCG 120  
TGGCGCTTTCTGCGTATCCTCTGCAAGACGGCGAGGCGAG 160  
ACCTCTTTGGCCTCTCTGTCTGATTCTGTGTTCCGCTAGA 200  
210 220 230 240  
GCTGCGACGACACCGGGCGAGCAGGAGACACGATCCCGTGC 240  
ATCTTCCAGGCTGTGGCCCGGGCGACAACCAGAGCGCTGG 280  
CACTGGTGGACGCGCAGTAGTGGTATATGCTGGACCTTCGC 320  
ACAGCTGGACACCTACTCCAAATGCTGTAGCCAACTGTTG 360  
CGCCAGCTGGGCTTTGCAACCAGGCGATGTGGTGGCTGTGT 400  
410 420 430 440  
TCCTGGAGGGCGCGCGGGAGTTCTGTGGGACTGTGGCTGGG 440  
CCTGGCCAAAGGCGGGTGTGGTGGCTGCTCTTCTCAATGTC 480  
AACCTGAGGGCGGGAGGCCCTGGCCTTCTGCTGGGCACAT 520  
CAGCTGCCAAAGGCCCTCATTTATGGCGGGGAGATGGCAGC 560  
GGCGGTGGGGGAGGTGAGCGAGCAGCTGGGGGAAGAGCCTC 600  
610 620 630 640  
CTCAAGTTCTGCTCTGGAATCTGGGGCCTGAGAGCATCC 640  
TGCCTGACAGGCGAGCTGCTGGACCCCATGCTTGTGTAGGC 680  
GCCCAGCAGACCCCTGGCACAAGCCCCAGGCAAGGGCATG 720  
GATGATCGGCTGTCTTACATCTATACTTCTGGGACCAACC 760  
GGCTTCTTAAGGCTGCCAATTGTGGTGCACAGCAGGTACTA 800  
810 820 830 840  
CCGCATTGCTGCTTTTGGCCACCATTCCTACAGCATGCGT 840  
GCCGCGGATGTGCTCTATGACTGCTGCGCACTCTACCACT 880  
CTGCAGGGGAACATCATGGGTGTGGGGCAGTGCCTCATCTA 920  
CGGGTTGACGGTGGTACTGCGCAAGGAAGTTCTCCGCCAGC 960  
CGCTTCTGGGATGACTGTGTCAAGTACAATTGCACGGTAG 1000  
1010 1020 1030 1040  
TGGATGACATAGGTGAATCTGCGCGCTACCTGCTGAGGCA 1040  
GCCGGTTCCGACGTGGAGCAGCGACACCGCGTGGCGCTG 1080  
GCCGTGGGTAAATGGGCTGGCGGCCAGCCATCTGGGAGGAGT 1120  
TCACGCAGCGCTTCGGTGTGCCACAGATCGGCGAGTTCTA 1160  
CGGCGCTACCGAGTGCACACTGCAGCATTGCCAACATGGAC 1200

Fig. 62A



1210 1220 1230 1240  
GGCAAGGTGGGCTCCTGCGGCTTCAACAGCCGTATCCTCA 1240  
CGCATGTGTACCCCATCCGTCTGGTCAAGGTCAATGAGGA 1250  
CACGATGGAGCCACTGCGGGACTCCGAGGGCCTCTGCATC 1320  
CCGTGCCAGCCCGGGGAACCCGGCCTTCTCGTGGGCCAGA 1360  
TCAACCAGCAGGACCCCTCTGCGGCGTTTCGATGGTTATGT 1400  
1410 1420 1430 1440  
TAGTGACAGTGCCACCAACAAGAAGATTGCCACAGCGTT 1440  
TTCCGAAAGGGCCGATAGCGCCTACCTCTCAGGTGACGTGC 1460  
TAGTGATGGAGCTGGGCTACATGTATTTCCGTGACCG 1520  
CAGCGGGGACACCTTCCGCTGGCGCGGGGAGAACGTGTCC 1560  
ACCACGGAGGTGGAAGCCGTGCTGAGCCGCGCTACTGGGCC 1600  
1610 1620 1630 1640  
AGACGGACGTGGCTGTGTATGGGGTGGCTGTGCCAGGAGT 1640  
GGAGGGGAAGCTGGCATGGCAGCCATCGCAGATCCCCAC 1680  
AGCCAGTTGGACCCCTAACTCAATGTACCAGGAATTACAGA 1720  
AGGTTCTTGCATCCTATGCTCGGCCCATCTTCTCTGCGTCT 1760  
TCTGCCCCAGGTGGATACCCACAGGCACCTTCAAGATCCAG 1800  
1810 1820 1830 1840  
AAGACCCGGCTGCAGCGTGAAGGCTTTGACCCGCCCTCAGA 1840  
CCTCAGACAGGCTCTTTCTTTCTAGACCTGAAGTCCGGCAC 1880  
GAGGTATCTACCCCTGGATGAGAGAGTCCATGCCCGCATT 1920  
TGCGCAGGGCAGCTTCTCACTCTGAGCCTGGAGAGTGGGCT 1960  
GGGCTTGGACTCCTGAGACCTGGGAGCCTGACACCCCTCT 2000  
2010 2020 2030 2040  
TCGGGTGCTTCTCTCTGCTTGGCCAGCATGGACAGCAGCACC 2040  
TGTGAGAGTAGGAAAAATGGAACCTGAGTGGCTGGGACCCC 2080  
TCTCTACTTCCCACTATGCATCCATTTTGGCTCTGCTTT 2120  
GATCTTTTTCTCCATCTCTTTTCTCCCTACCCAGCAGGAG 2160  
CCCCACAAACACATGTTGGCTGCTGTGTCTCTGAGTTGGA 2200  
2210 2220 2230 2240  
CCAGTGTCCAGGGGTACAGGCTTCAGGCTGTGACCCACAC 2240  
TGGTACCCACCTCCCTTTTCTATTTTGGCTTAGGTTTCATC 2280  
CACGGTTCCCTGTGGAGCAAGTGGGGGCCACATAGCTG 2320  
CTGTCCCTGCTGAGGGTTGGTAGCAATCACACCCCTCATGT 2360  
CAGCTGGGAGACACGCGCAGTCTGCCACTGACCCCAATC 2400  
2410 2420 2430 2440  
AACTGAAAAATATTGTTTTGACTACTTTTTGTTTTTTGTT 2440  
TTTTTGTTTTTTTTTTTTTTTCGAGACAGAGTTTCTCTGTA 2480  
TAGCCCTGGCTGTCTCTGGAACCTCACTTTGTAGACCAAGCT 2520  
GGCCTCGAACTCAAAAAATCCTCCTGACTCTGCCTCTGCTT 2560  
CCCAAGTGCTGGGATTAAAGACGTGGCGCCACCACCGCCTG 2600

Fig. 62B

```
      2610      2620      2630      2640
.....
GCTGTTTTGTATTTTGTGTTTGTGTTTGACGATAGGGTCTC 2640
ACTGTGGAGGCCAAGCTGGCCTCAGACTCCCCACCCATT 2660
GCCTCTGGGCACCATTTCTATATTCTCAGACTGATGACAAT 2720
GCACTAGTGTCCCTAGGAGTCTTGAGTCTGCACTTTCCCC 2760
TCATAGCCTCAAGCTTCCAGAACTGACTCTGATCACTTGG 2800

      2810      2820      2830      2840
.....
ATGTGGCTAGTGTGTTGGCTCTACCCACATGTGTCAATTCAG 2840
GGGTCCCCAGGCATAGTCTCTGGAAGCCCTCACCCGGAAA 2880
AAGCTTGGAGAGACCCAGGAAGGTTGTTGTGTTCTCTTGG 2920
GCACCCCTGGTGGCAGTCCTGGGCATGCTTCCGCACTGT 2960
ACTGGTGCATATAGCCAGACCTATGACATTTGAGGTCTA 3000

      3010      3020      3030      3040
.....
CCCTTCTGGCTCCTGTGTTGCCCAATTGAGATCCTTGGTGA 3040
CTCACCTCAGTCACCAAGCAGAGCCTCTGCCTGCCTTCAT 3080
CTTCAAGGTCATGAAGGATGTGGACAGAGCACTACAGGC 3120
TGCCAGCAGTCAACCACATGAGAGTGTTACTTCCTTGTTG 3160
GTTTTTAAAAAATAAAATGTGCTGAGCCTCGAAAAAAAAAA 3200

      3210      3220      3230      3240
.....
AAAAAAAAAAAAAAAAAA 3217
```

Fig. 62C

10	20	30	40
MRAPGAGTASVASLALLWFLGLPWTWSAAAAFCVYVGGGG	40		
WRFLRIVCKTARRDLFGLSVLIRVRLRLRRHRRAGDTIPC	80		
IFCAVARRQPERLALVDASSGICWTFAGLDTYSNAYANLF	120		
RCLGFAPGQVYAVFLEGRPEFVGLWLGLAKAGVVAALLNV	160		
NLRREPLAFCLGTSAAKALIYGGEMAAAVAENVSECLGKSL	200		
210	220	230	240
LKFCSGDLGPFESILFOTQLLOFMLAEAPTTFLACAPGKGM	240		
DDRIFYIYTSGETTGLPKAAIVVHSRYRRAAFQHHSYSMR	280		
AADVLYDCLPLYHSAGNIMGVGGCVIYGLTVVLRKKFSAS	320		
RFWDCQVKYNCTVYDDIGETCRYLLRGFVROVEQRHRVRL	360		
AVGNGLRFAIWEFTGRFGVPGIGEFYGATECNCSIANMD	400		
410	420	430	440
GKVGSGGPNRILTHVYFIRLVKVNEDTMEPLRDSGLCI	440		
PCGPGEGQLLVGGINCCGFLRRFGGYVSGSATNKKIAHSV	480		
FRKGSAYLSSGVLYMDELGYMYFRDRSGGTFRWGQENVB	520		
TTEVEAVLSRLLGQTOYAVYGVAVFGVECKAGMAAIAQPH	560		
SGLDPNMSMYGELQKVLASYARFIFLRLLPGVDTTGTGFKIC	600		
610	620	630	640
KTRLCRSGFGPFGTSGRLFFLDLKGSTRYLFLQSRVHAR	640		
CAGQFSL	647		

Fig. 63

10 20 30 40  
GGGCGGAGGCGGAGGCCAGTCGCCAGCTCCTGCTCTGCTC 40  
CTCTCCCGGCTGCGCGCGCGCTGCACGCCCTCGAGCACTCC 80  
CTCGGCCCCCGGCGGGGACCGGGGACCCCGCAGCTACCGCC 120  
ATGCTGCCAGTGTCTCTACACCGGCTTGGCGGGGCTGCTGC 160  
TGCTGCCCTCTGCTGCTCACCTGCTGCTGCCCTACCTCCT 200  
210 220 230 240  
CCAAGATGTGCGGTACTTCTGCGGCTGGCCAAACATGGCC 240  
CGGCGGGTGGCGAGCTACCGGCAGCGGCGACCTGTGCTA 280  
CCATCCTGCGGGCTTCTTGGAAACAAGCGCGCAAGACCCC 320  
ACACAAGCCCTTCTGCTGTTCCGAGACGAGACGCTCACC 360  
TACGCCCAGGTGGACCGGCGCAGCAACCAAGTGGCGCGGG 400  
410 420 430 440  
CGCTGCACGATCAACTGCGGCTACGACAGGGGGATTGCGT 440  
AGCCCTCTTCATGGGCAATGAGCGCGCTACGTGTGGATC 480  
TGGCTGGGACTGCTCAAACTGGGCTGTCCCATGGCGTCC 520  
TCAACTACAACATTGCTGCCAAGTCTCTGCTGCAGTCTT 560  
TCAATGCTGCGGGCGAAGGTGCTGCTGGCTCCCCAGAT 600  
610 620 630 640  
CTACAAGAAGCTGTGGAGGAGGTTCTTCCAAACCTGAAAA 640  
AGGATGGCGGTGTCCGTCTTTTACGTAAAGCAAACTTCTAA 680  
CACAAATGGGTGTGGACACAATACTGGACAAAAGTAGACGGA 720  
GTGTGGGCGGAACCCACCCCGGAGTGTGGAGGTGTGAAG 760  
TCACTTTTACCACGCCAGCAGTATACATTTTACTTTCGGG 800  
810 820 830 840  
AACCACAGGTCTTCCAAAAAGCGGAACCATCAATCATCAT 840  
CGCCTAAGGTATGGGACAAGCCTTGCTATGTGAGTGGGA 880  
ATCACGGGCAAGGATGTCACTATATACCAACAATGCCCTTG 920  
JTCCAAACAGTGCACCGCTCAAGATCGGCTTTCAGCGATGC 960  
ATCCTGGGTGTGGGCTACTTTTAACTTGGCGGGGCAAAAT 1000  
1010 1020 1030 1040  
CTCAAGCAAGCCAAATTTTGGGAACGACTGGCAGGAAATAC 1040  
AACGTCAACGGTCAATTCAGTACATTGGTGAACCTGCTTCGG 1080  
TACCTGTGCAACACACCGCAGAAACCAAAATGACCGGGACC 1120  
ACAAAGTGAAAAAAGCCCTGGGAAATGGCTTACGAGGAGA 1160  
TGTTGGAGAGAGTTTCAATCAAGAGATTGGGGACATCCAC 1200

Fig. 64A

1210	1220	1230	1240
GTGTATGAGTTCTACGCATCCACTGAAGGCAACATTGGAT	1240		
TTGTGAACATATCCAAGGAAAAATCGGTGCTGTGGGAGAGC	1250		
AAACTACCTACAAAGAAAAAGTTGCAAGGTATGAGCTGATC	1320		
AAGTATGACGTGGAGAAGGACGAGCCGGTCCGTGACGCAA	1360		
ATGATATTGCATCAAAGTCCCCAAAGGTGAGGTTGGACT	1400		
1410	1420	1430	1440
CTTGGTTTGC AAAATCACACAGCTCACACCATTATTGGC	1440		
TATGCTGGAGGAAAGACCCAGACAGAGAAGAAAAAACTCA	1480		
GAGATGTCTTTAAGAAAGGGGACATCTACTTCAACAGCGG	1520		
AGACCTCCTGATGATCGACCGTGAGAACTTCGTCTACTTT	1560		
CACGACAGGGTTGGAGATACTTTCCGGTGGAAAGGAGAGA	1600		
1610	1620	1630	1640
ACGTAGCTACCACAGAAGTGGCTGACATCGTGGGACTGGT	1640		
AGATTTTGTGAAGAAGTGAATGTGTATGGCGTGCCTGTG	1680		
CCAGGTCATGAGGGTGAATTTGGCATGGCCTCCCTCAAGA	1720		
TCAAAGAAAACTACGAGTTCAATGGAAAGAAACTCTTTCA	1760		
ACACATCGCGGAGTACCTGCCCCAGTTACGGGAGGCTCGG	1800		
1810	1820	1830	1840
TTCTTGAGGATACAAAGATACCATTTGAGATCACTGGGACTT	1840		
TTAAACACCGCAAGTGACCCCTGATGGAAGAGGGCTTCAA	1880		
TCCCACAGTCATCAAAGATACCTTGTATTTTCATGGATGAT	1920		
GCAGAGAAAAACATTTGTGCCCCATGACTGAGAACATTTATA	1960		
ATGCCATAAATTGATAAAAACCTCTGAAGCTCTGAATATTCCC	2000		
2010	2020	2030	2040
TGGTGGTTTACGTCATGACATTTTCAGAAAAAGAACTGGAT	2040		
AGACCTCGCAGAGGCCACTTCATATGTAGAAATCCAACTTTA	2080		
ACTTGATTGAAGACTATAAGGTGGGATTTTATTTTATAGGA	2120		
AATTATTGATTAAAGGATAGTTTTTTTTTTTTTTTTTTAA	2160		
TTACACCTGAACCTTTGCAAGTAAAAAGATTTAAGACAAA	2200		
2210	2220	2230	2240
TTATTTTTTCAATGTGCACCTGCCATTTGTCTTTGC AAACT	2240		
AAGCTTCTTGGAGAGAGGGCCTTATTTTTTTTAAAGACATA	2280		
ATAACTATATTAACTATAAAAAAAAAAAAAAAAAAAAAA	2320		
AAAAAAAAAAAAAAAAAAAAA	2360		

Fig. 64B

10	20	30	40
MLFVLYTGLAGLLLLPLLLTCCCPYLLQGVRYFLRLANMA	40		
RRVRSYRCRRFVRTILRAFLEGARKTPHKPFLLFRDETLT	60		
YACVGRRSNCGVARALHDGLGLRCGDCVALFMGNEPAYVW	120		
WLGLLKLGCPMACLNYNIRAKSLLHCFQCCGAKVLLASPD	160		
LGEAVEEVLFPTLKKDAVSVFYVSRTSNTNGVDTILDKVGG	200		
210	220	230	240
VSAEPTFESWRSEVTFITPAVYIYTSGTTGLPKSGTINHH	240		
RLRYGTSLAMSSGNHGGGCHLYCQCPCSNSATLKIGLHGC	280		
ILGWGYFNLGGANSCASCFWERLAGNTTSTVIGYIGELLR	320		
YLNTFCKFNDRCHKVKKALGNGLRGQVWREFIKRFGQIH	360		
VYEFYASTEIGNIGFVNYPRKIGAVGRANYLGRKVARVEL	400		
410	420	430	440
KYQVEKQDFVRDANGYCIKVPKGEVGLLVCKITCLTFPIG	440		
YAGGKTCTEKKKLROVFKKGGIYFNSGGLLYDRENFTYF	480		
HCRVGGTFRWKGENVATTEVADIVGLVDFVEEVNYYGVPY	520		
PGHEGRIGMASLKIKENYEFNGKKLFCHIAEYLPYARPR	560		
FLRIQCTIEITGTFKHKVTLMEEGFNFTVVKDTLYFMDQ	600		
610	620	630	640
AEKTFVPMTENIYNALIDKTLKL	623		

Fig. 65

10 20 30 40  
GAAAGCTCTGAGAGCGGGTGCAGTCTGGCCTGGCGTCTCG 40  
CGTACCTGGCCCGGGAGCAGCCGACACACACCTTCCTCAT 80  
CCACGGCGCGCAGCGCTTTAGCTACGCGGAGSCTGAGCGC 120  
GAGAGCAACCGGATTGCTCGCGCCTTTCTGCGCGCACGGG 160  
GCTGGACCGGGGGCGCCGAGGCTCGGGCAGGGGCAGCAC 200  
210 220 230 240  
TGAGGAAGGGCGCACGCGTGGCGCCTCCGGCTGGAGATGCG 240  
GCTGCTAGAGGGACGACCGCGCCCCCTCTGSCACCGGGG 280  
CGACCGTGGCGCTGCTCCTCCCAGCGGGCGCGGATTTCCT 320  
TTGGATTTGGTTGCGACTGGCCAAAGCTGGCTGCGCACG 360  
GCTTTGTGCCCCACCGCTTTACGCCGAGGACCCCTGCTGC 400  
410 420 430 440  
ACTGCTCCGCGAGCTGCGGTGCGAGTGCCTGCTGCTGGC 440  
CACAGAGTTCTCTGGAGTCCCTGGAGCCCGGACCTGCCGSC 480  
TTGAGAGCCATGGGGCTCCACCTATGGGCGACGGGCCCTG 520  
AAACTAATGTAGCTGGAAATCAGCAATTTGCTATCGGAAGC 560  
AGCAGACCAAGTGGATGAGGCCAGTGGCGGGGTACCTCTCT 600  
610 620 630 640  
GCCCCCAGAAACATAATGGACACCTGCTCTACATCTTCA 640  
CCTCTGGGACTACTGGCTGCCCCAAGGCTGCTGGAAATCAG 680  
TCACTGAAAGGTTCTACAGTGGCAGGGATTTTACCATCTG 720  
TGTGGAGTCCACAGGAGGAGCGTGATCTACCTGGCACTGC 760  
CACTGTACCACATGTCTGGCTCCCTTCTGGGCAATTGT 800  
810 820 830 840  
CTGCTTGGGCATTGGGGCCACCGTGGGTGCTGAAACCCCAAG 840  
TTCTCAGCTAGCCAGTTCTGGGACGATTGCCAGAAACACA 880  
GGGTGACAGTGTTCCAGTACATTGGGGAGTTGTGCCGATA 920  
CCTGGTCAATCAGCCCGCGAGCAAGGCAGAGTTTGACCAT 960  
AAGGTGCGCTTGGCAGTGGGCAGTGGGTGCGGCCAGACA 1000  
1010 1020 1030 1040  
CCTGGGAGCGTTTCTCTGCGGCGATTGGAACCTCTGCAGAT 1040  
ACTGGAGACGTATGGCATGACAGAGGGCAACGTAGCTACG 1080  
TTCAATTACACAGGACGGCAGGGTGCAGTGGGGCGAGCTT 1120  
CCTGGCTTTTACAAGCACATCTTCCCTTCTGCTTGATTGG 1160  
ATACGATGTGATGACAGGGGAGCCTATTGGGAATGCCCAAG 1200

Fig. 66A

1210	1220	1230	1240
GGGCACTGCATGACCACATCTCCAGGTGAGCCAGGCCTAC	1240		
TGGTGGCCCCAGTGAGCCAGCAGTCCCCCTTCCTGGGCTA	1280		
TGCTGGGGCTCCGGAGCTGGCCAAGGACAAGCTGCTGAAG	1320		
GATGTCTTCTGGTCTGGGGACGTTTTCTTCAATACTGGGG	1360		
ACCTCTTGGTCTGTGATGAGCAAGGCTTTCTTCACTTCCA	1400		

1410	1420	1430	1440
CGATCGTACTGGAGACACCATCAGGTGGAAGGGAGAGAAT	1440		
GTGGCCACAACCTGAAGTGGCTGAGGTCTTGGAGACCCCTGG	1480		
ACTTCCTTCAGGAGGTGAACATCTATGGAGTCACGGTGCC	1520		
AGGGCAGGAAGGCAGGGCAGGCATGGCGGCCCTTGGCTCTG	1560		
CGGCCCCCGCAGGCTCTGAACCTGGTGCAGCTCTACAGCC	1600		

1610	1620	1630	1640
ATGTTTCTGAGAACTTGCCACCGTATGCCCGACCTCGGTT	1640		
TCTCAGGCTCCAGGAATCTTTGGCCACTACTGAGACCTTC	1680		
AAACAGCAGAAGGTTAGGATGGCCAATGAGGGCTTTGACC	1720		
CCAGTGTACTGTCTGACCCACTCTATGTTCTGGACCAAGA	1760		
TATAGGGGSCCTACCTGCCCTCACACCTGCCCGGTACAGT	1800		

1810	1820	1830	1840
GCCCTCCTGTCTGGAGACCTTCGAATCTGAAACCTTCCAC	1840		
TTGAGGGAGGGGCTCGGAGGGTACAGGCCACCATGGCTGC	1880		
ACCAGGGAGGGTTTTTCGGGTATCTTTTGTATATGGAGTCA	1920		
TTATTTTGTAAATAAACAGCTGGAGCTTAAAAAAAAAAAAA	1960		
AA	1996		

Fig. 64B



10	20	30	40
ESSESGLAWRLAYLAREQPTHTFLIHGACRFSYAEAE	40		
ESNRIRAFRLRARGWTGGRRGSGRGSTEEGARYVAPPAGDA	80		
AARGTTAPFLAPGATVALLLPAGPOFLWIWFGGLAKAGLRT	120		
AFVPTALRRGPLLHCLRSCGASALVLATEFLESLEFOLPA	160		
LRAMGLHLWATGPFETNVAGISNLLSEAAOQVDEPVFGYLS	200		
210	220	230	240
APCNIMDTCLYIFTSGTTGLPKAARTSHLKVLCGGGFYHL	240		
CGVHQEDVIYLALFLYHMSGSLLGIVGCLGIGATVVLKPK	280		
FSASGFWDCCCKHRVTVFQYIGELCRYLVNCFPSKAEFOH	320		
KVRLAVSGSLRPOTWERFLRRFGPLQILETYGMEGNVAT	360		
FNYTGRCCGAVGRASWLYKHIFFPSLIRYQVMTGEPINAG	400		
410	420	430	440
GHCMTTSPGEPGLLVAPVSGGSPFLGYAGAPELAKDKLLX	440		
GVFWSGGVFFNTGOLLVQDEGGFLHFFHRTGOTIRWKGEN	480		
VATTEVAEYLETLOFLCEVNIYGVTVFGHEGRAGMAALAL	520		
RPPCALNLVGLYSHVSENLPFYARPRFLRLCESLATTETF	560		
KGGKVRMANEGFOPSVLSOPLYVLOGGIGAYLPLTPARYS	600		
610	620	630	640
ALLSGDLRI	609		

Fig. 67

```
      10      20      30      40
.....
ATGCTGCTTGGAGCCTCTCTGGTGGGGGCGCTACTGTTCT 40
CCAAGCTAGTGTCTGAAGCTGCCCTGGACCCAGGTGGGATT 80
CTCCCTGTTGCTCCTGTACTTGGGGTCTGGTGGCTGGCGT 120
TTCATCCGGGTCTTTCATCAAGACGGTCAGGAGAGATATCT 160
TTGGTGGCATGGTGTCTCTGAAGGTGAAGACCAAGGTGCG 200

      210      220      230      240
.....
ACGGTACCTTCAGGAGCGGAAGACGGTGGCCCTGCTGTTT 240
GCTTCAATGGTACAGCGCCACCCGGACAAAGACAGGCTGA 280
TTTTCGAGGGGCACAGACACTCACTGGACCTTCCGCGAGCT 320
GGATGAGTACTCCAGTAGTGTGGCCAACTTCCCTGCAGGCC 360
CGGGGCTGSCCTCAGGCCAATGTAGTTGCCCTCTTTATGG 400

      410      420      430      440
.....
AAAACCGCAATGAGTTTGTGGGTCTGTGGGTAGGCATGGC 440
CAAGCTGGGCGTGGAGGCGGGCTCTCATCAACACCAACCTT 480
AGGGGGGATGCCCTGCGCCACTGTCTTGACACTTCAAAGG 520
CAAGAGCTCTCATCTTTTGGCAGTGAGATGGCCTCAGCTAT 560
CTGTGAGATCCATGCTAGCCTGGAGGCCACACTCAGCCTC 600

      610      620      630      640
.....
TTCTGCTCTGGATCCTGGGAGCCCGAGCACAGTGCCTGTC 640
GCACAGAGCATCTGGACCTCTTCTGGAGATGCCCGGAA 680
GCACCTGGCCAGTCAACCCAGACAAAGGTTTTCAGATAA 720
CTCTTCTACATCTACACATCGGGCAACAGGGGGTACCCA 760
AAGGTGGCATTGTGGTGCACAGCAGGTATTATCTATGGC 800

      810      820      830      840
.....
TTCCCTGGTGTACTATGGATTCCGCATGGGGCCTGATGAC 840
ATTGTCTATGACTGCCCTCCCGCTCTACCACTCAAGCAGGA 880
AATCTGGTGGGGATTGGCAGTGGTTACTCCACGGCATGAC 920
TGTGGTGAACCGGAAGAAGTTCTCAGCCTCCCGGTTCTGG 960
GATGATTGTATCAAGTACAACTGCACAGTGGTACAGTACA 1000

      1010      1020      1030      1040
.....
TTGGCGAGCTCTGCGGCTACCTCCTGAACGAGCCACCCCG 1040
TGAGGCTGAGTCTCGGCACAAAGGTGCCCATGGCACTGGGC 1080
AACGGTCTCCGGCAGTCCATCTGGACCGGACTTCTCCAGCC 1120
GTTTCCACATCCCCCAGGTGGCTGAGTTCTATGGGGCCAC 1160
TGAATGCCAACTGTAGCCTGGGCAACTTTGACAGCCGGGTG 1200
```

Fig. 68A

1210	1220	1230	1240
GGGGCCTGTGGCTTCAATAGCCGCATCCTGTCTTTGTGT	1240		
ACCCATATCCGTTTGGTACGTGTCAATGAGGATACCATGGA	1280		
ACTGATCCCGGGGACCCGATGGAGTCTGCATTCCCTGTCAA	1320		
CCAGGTCAGCCAGGCCAGCTGGTGGGTGCGATCATCCAGC	1360		
AGGACCCCTCTGCGCCGTTTCGACGGGTACCTCAACCAGGG	1400		
1410	1420	1430	1440
TGCCAACAACAAGAAGATTGCTAATGATGTCTTCAAGAAG	1440		
GGGGACCAAGCCTACCTCACTGGTGACGTCTTGGTGATGG	1480		
ATGAGCTGGGTTACCTGTACTTCCGAGATCGCACTGGGGA	1520		
CACGTTCCGCTGGAAAAGGGGAGAATGTATCTACCACTGAG	1560		
GTGGAGGGGCACACTCAGCCGCCCTGCTTCATATGGCAGATG	1600		
1610	1620	1630	1640
TGGCAGTTTATGGTGTGTGAGGTGCCAGGAAGTGAAGGCGG	1640		
AGCAGGAATGGCTGGCGTTGCAAGTCCCATCAGCAACTGT	1680		
GACCTGGAGAGCTTTGCACAGACCTTGAAAAAGGAGCTGC	1720		
CTCTGTATGCCCGCCCCATCTTCTCTGCGCTTCTTGGCTGA	1760		
GCTGCACAAGACAGGGACCTTCAAGTTCCAGAAGACAGAG	1800		
1810	1820	1830	1840
TTGCGGAAGGAGGGCTTTGACCCATCTGTTGTGAAAGACC	1840		
CGCTGTTCTATCTGGATGCTGGGAAGGGCTGCTACGTTGC	1880		
ACTGGACCAGGAGGGCTATACCCGCACTCCAGGCAGGGCAG	1920		
GAGAAGCTGTGATTTCCCGCTACATCCCTCTGAGGGCCAG	1960		
AAGATGCTGGATTTCAGAGGCCCTAGCGTCCACCCAGAGGG	2000		
2010	2020	2030	2040
TCCTGGGCAATGCCAGAGCAAAAGCTAGCAGGGCCCGCACC	2040		
TCCGCCCCCTAGGTCGTGATCTCCCTCTCCCAAACTGCCA	2080		
AGTGACTCACTGCCGCTTCCCGCACTCTCCAGAGGCTTTTC	2120		
TGTGAAAGTCTCATCCAAAGCTGTGTCTTCTGGTCCAGGCG	2160		
TGGCCCCCTGGCCCCAGGGTTTCTGATAGGCTCTTTAGGA	2200		
2210	2220	2230	2240
TGGTATCTTGGGTCCAGCGGGGCCAGGGGTGTGGGAGAGGAG	2240		
TCACTAAGATCCCTCCAAACAGAAAGGGAGCTTACAAAGGA	2280		
ACCAAGGCCAAAAGCCTGTAGACTCAGGAAGCTAAGTGGCCA	2320		
GAGACTATAGTGGCCAGTGAATCCCATGTCCACAGAGGATC	2360		
TTGGTCCAGAGCTGCCAAAGTGTACCTCTCCCTGCCCTGC	2400		
2410	2420	2430	2440
ACCTCTGGGGAAAAAGAGGACAGCATGTGGCCACTGGGCAC	2440		
CTGTCTCAAGAAAGTCAGGATCACACACTCAGTCTCTTGT	2480		
CTCCAGGTTCCCTTGTCTTGTCTCGGGGAGGGAGGGAGC	2520		
AGTGTCTGTCTGTCTCTTCTGCTGTCTGTGAGTCTGTG	2560		
TTGCTTCTCCATCTGTCTTAGCCTGAGTGTGGGTGGAAACA	2600		

Fig. 68B

2610 2620 2630 2640  
.....  
GGCATGAGGAGAGTGTGGCTCAGGGGCAATAAACTCTGC 2640  
CTTGACTCCTCTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2680  
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2710

Fig. 68C

10	20	30	40
MLLGASLVGALLFSKLVLKLPWTVGFSLULLYLGSGGWR	40		
FIRVFIKTVRRGIFCGMVLLKVKTKVRRYLGERKTVPLLF	80		
ASMVQRHFOKTALIFEGTOTHWTFRQLDEYSSSVANFLCA	120		
RGLASGNVVALFMENRNEFVGLWLGMAKLGVEAALINTNL	160		
RRDALRHCLDTSKARALIFGSEMASAICEIHASLEPTLSL	200		
210	220	230	240
FCSGSWEPSTVPVSTZHLQPLLEDAFKHLPSHPCKGFTBK	240		
LFYIYTSGETTGLPKAAIVVHSRYRMAASLVYYGFRMRPDD	280		
IYYOCLFLYHSSRKHRGGWCCLLHGMTVVIRKKFSASRFW	320		
DDCIKYNCTVVQYIGELCRYLLNQPFRFAESRHKYRMALG	360		
NGLRQSIWTFSSRFHIFQVAEFYGATECNCSLGNFDSRV	400		
410	420	430	440
GACGFNSRILSFVYFIRLVRVNEDFMELIRGPDGVCIPOI	440		
PGGPGCLVGRITGGDFLRFDGYLNGGANMKKIANGVFKK	480		
GCGAYLTGGVLVMGELGYLYFRDRTGOTFRWKSENVSTTE	520		
VEGTLRLLHMAQVAVYGVENVPGTEGRAGMAVVASPISNC	560		
DLSSFACTLKKEFLYARPIFLRFLPELHKTGTTFKPKTE	600		
610	620	630	640
LRKSGFQPSVYKQFLFYLDARKSCYVALDGEAYTRICAGE	640		
EKL	643		

Fig. 69

10 20 30 40  
CACTCATCAGAGCTAAGAGAGACTACACGCTCTCATCTAC 40  
TTCAGAAAGAGCCAATGCCATGGGTATTTGGAAGAAACTA 80  
ACCTTACTGCTGTTGCTGCTTCTGCTGGTTGGCCTGGGGC 120  
AGCCCCCATGGCCAGCAGCTATGGCTCTGGCCCTGCGTTG 160  
GTTCTGCGGAGACCCACATGCCTTGCTGCTGCTTGGCTTG 200  
210 220 230 240  
GCATTGCTGGGCAGACCCCTGGATCAGCTCCTGGATGCCCC 240  
ACTGGCTGAGCCTGGTAGGAGCAGCTCTTACCTTATTCT 280  
ATTGCTCTACAGCCACCCCCAGGGCTACGCTGGCTGCAT 320  
AAAGATGTGGCTTTTACCTTCAAGATGCTTTTCTATGGCC 360  
TAAAGTTGAGGCGAGCCCTTAACAAACATCCTCCAGAGAC 400  
410 420 430 440  
CTTTGTGGATGCTTTAGAGCGGCAAGCACTGGCATGGCCT 440  
GACCGGGTGGCCTTGGTGTGTACTGGGTCTGAGGGCTCCT 480  
CAATCACAATAAGCCAGCTGGATGCCAGGTCTGTGAGGC 520  
AGCATGGGTCTGAAAGCAAGCTGAAGGATGCCGTAAATC 560  
CAGAACACAAGAGATGCTGCTGCTATCTTAGTTCTCCCGT 600  
610 620 630 640  
CCAAGACCATTTCTGCTTTGAGTGTGTTTCTGGGGTTGGC 640  
CAAGTTGGGCTGCCCCGTGGGCTGGATCAATCCACACAGC 680  
CGAGGGATGCCCTTGTACACTCTGTACGGAGCTCTGGGG 720  
CCAGTGTGCTGATTGTGGATCCAGACCTCCAGGAGAACCT 760  
GGAAGAGTCTCTCCCAAGCTGCTAGCTGAGAGCATTGAC 800  
810 820 830 840  
TGCTTCTACCTTTGGCCACAGCTCACCCACCCCGGGAGTAG 840  
AGGCTCTGGGAGCTTCCCTGGATGCTGCACCTTCTGACCC 880  
AGTACCTGCCAGCCTTCCAGCTACGATTAAAGTGGAAATCT 920  
CCTGCCATATTCACTCTTTACTTTCAGGGACCACTGCACTCC 960  
CAAGGCCAGCCATCTTATCAGATGAGCGGGTCAATACAGT 1000  
1010 1020 1030 1040  
GAGCAACGTGCTGTCTTCTGTGGATGCAGAGCTGATGAT 1040  
GTGGTCTATGACGTCTTACCTCTGTACCATACGATAGGGC 1080  
TTGTCTTTGGATTCTTTGGCTGCTTACAAGTTGGAGCCAC 1120  
CTGTGTCTTGGCCCCCAAGTTCTCTGCTCCCGATTCTGG 1160  
GCTGAGTGCCTGGCAGCATGCCGTAACAGTGATCTTGTATG 1200

Fig. 7cA

1210	1220	1230	1240
TGGGTGAAATCCTGCGGTACTTGTGTAAACGTCCCTGAGCA 1240			
ACCAGAAGACAAGATACATACAGTCCGCTTGGCCATGGGA 1280			
ACTGGACTTCGGGCAAATGTGTGGAAAACTTCCAGCAAC 1320			
GCTTTGGTCCCATTCCGATCTGGGAATTCTACGGATCCAC 1360			
AGAGGGCAATGTGGGCTTAATGAACTATGTGGGCCACTGC 1400			
1410	1420	1430	1440
GGGGCTGTGGGAAGGACCAGCTGCATCCTTCCAATGCTGA 1440			
CTCCCTTTGAGCTTGTACAGTTCCGACATAGAGACAGCAGA 1480			
GCCTCTGAGGGACAAACAGGGTTTTTGCATTCTGTGGAG 1520			
CCAGGAAAAGCCAGGACTTCTTTTGACCAAGGTTGGAAGA 1560			
ACCAACCCTTCTGGGCTACCGTGGTTCCAGGCCGAGTC 1600			
1610	1620	1630	1640
CAATCGGAAACTTGTTCGGAATGTACGACGGGTAGGAGAC 1640			
CTGTACTTCAACACTGGGGACGTGCTGACCTTGGACCAGG 1680			
AAGGCTTCTTCTACTTTCAAGACCGCCTTGGTGACACCTT 1720			
CCGGTGGAAAGGGCGAAAACGTATCTACTGGAGAGGTGGAG 1760			
TGTGTTTTGTCTAGCCTAGACTTCTTAGAGGAAGTCAATG 1800			
1810	1820	1830	1840
TCTATGGTGTGCTGTGCCCAGGGTGTGAGGGTAAGGTGG 1840			
CATGGCTGCTGTGAAAAGTGGCTGCTGGGAAGACTTTTGAT 1880			
GGGCAGAAAGCTATACCAGCATGTCCGCTCCTGGCTCCCTG 1920			
CCTATGCCACACCTCATTTTCATCCGTATCCAGGATTCCT 1960			
GGAGATCACAAAACCTACAAGCTGGTAAAGTCAGGGCTG 2000			
2010	2020	2030	2040
GTGCGTGAGGGTTTTGATGTGGGGATCATTGCTGACCCCC 2040			
TCTACATACTGGACAAACAAAGGCCAGACCTTCGGGATCT 2080			
GATGGCAGATGTGTACCAGGCTGTGTGTGAAGGAACCTGG 2120			
AATCTGTGACCACCTAGCCAACTGGAAAGGCAATCCAAAAG 2160			
TGTAGAGATTGACACTAGTCAGCTTCACAAAATTTGTCCGG 2200			
2210	2220	2230	2240
GTTCCAGATGCCCATGGCCCAAGTAGTACTTAGAGAATAAA 2240			
CTTGAATGTGTATACAAAAA 2277			

Fig. 70B

10 20 30 40  
MGIWKKLTLLLLLLLLLVGLGQFPWPAAMALALRWFLGQPT 40  
CLVLLGLALLGRFWISSWMPHWLSLVGAALTLLFLLPLQPF 80  
PGLRWLHKQVAFTFKMLFYGLKFRRRLNKHFFPETFVDALE 120  
RCALAWFORVALVCTGSESSITNSQLDARSQCAAWVLKA 160  
KLKDAVIGNTRQAAAIVLVFSKTIISALSVFLGLAKLGCPV 200  
210 220 230 240  
AWINFHSRGMPLLSVRSRGASVLIYDPLDGENLEEVLPK 240  
LLAENIHCFYLGHSSTFGVEALGASLOAAPSDPVFASLR 280  
ATIKWKSPAIFIFTSGTTGLPKPATLSHERVICVSNVLSF 320  
CGCRACQVYVDVLFPLYHTIGLVLGFLGCLQVQATCVLAPK 360  
FSASRFWAECRQHGVTVILYVGEILLRYLONVFEQPECKIH 400  
410 420 430 440  
TVRLAMGTGLRANVWKNFGCRFGFISWEFYGSSTEGNVGL 440  
MNYVGHQSAVGRTSCILRMLTFPELVQFOIETAEPLRCKI 480  
GFCIPVERPKPGLLLTKVRKNQFFLGVRGSAESNRKLVV 520  
NVRRAVGOLYFNTGOVLTLDGEGFFYFQDRLGQTFRWXGEN 560  
VSTGEVECVLSSLOFLEEVNVYGVYPGCEGKVGMMAAVKL 600  
610 620 630 640  
AFGKTFQGGKLYCHYRSWLPAYATPHFIRIQGSLEITNTY 640  
KLVKSRLVREGFQVGIADFLYILONKACTFRSLMFGVYQ 680  
AVCEGTWNL 689

Fig. 71



```

      10      20      30      40
.....
GCTCTCTGGGCCTATATCAAGCTGCTGAGGTACACGAAGC 40
GCCATGAGCGGCTCAACTACACGGTGGCGGACGTCTTCGA 80
ACGAAATGTTTCAGGCCCATCCGGACAAGGTGGCTGTGGTC 120
AGTGAGACGCAACGCTGGACCTTCCGTCAGGTGAACGAGC 160
ATCGGAACAAGGTGGCCAATGTGCTGCAGGCTCAGGGCTA 200
      210      220      230      240
.....
CAAAAAGGGCGATGTGGTGGCCCTGTTGCTGGAGAACCGC 240
GCCGAGTACGTGGCCACCTGGCTGGGTCTCTCCAAGATCG 280
GTGTGATCACACCGCTGATCAACACGAATCTGCGCGGTCC 320
CTCCCTGCTGCACAGCATCACGGTGGCCCATTTGCTCGGCT 360
CTCATTACGGCGAGGACTTCCTGGAAGCTGTCAACCGACG 400
      410      420      430      440
.....
TGGCCAAGGATCTGCCAGCGAACCCTCACACTCTTCAGTT 440
CAACAACGGAGAACAACAACAGCGAGACGGAAAAGAACATA 480
CCGCAGGCCCAAGAAATCTGAACGCGCTGCTGACCACGGCCA 520
GCTATGAGAAAGCCTAACAAGACCGCAGGTAAACCACCACGA 560
CAAGCTGGTCTACATCTACACCTCCGGCACCAACAGGABTG 600
      610      620      630      640
.....
CCAAAGGCTGGCGTTATCTCTCACTCCCGTTATCTGTTTA 640
TCGCTGCTGGCATCCACTACACCATGGGTTTTCCAGGAGGA 680
GGACATCTTTCTACACGGCCCTTGGCTTTGTACCAACCGCT 720
GGTGGCATTTATGGCATGGGTCAAGTGGGTGCTCTTTGGCT 760
CCACGGTCTCCATTCGCAAGAAAGTTCTCGGCATCCAACTA 800
      810      820      830      840
.....
TTTGGCGGACTGGCGCCAAGTATAATGCAACTATTGGTCAG 840
TATATCGGTGAGATGGCTCGCTACATTCTAGCTACGAAAC 880
CCTCGGAATACGACCAAGAAACACCGAGTGGGTCTGGTCTT 920
TGGAAACGGACTGGGACCGGCAGATTTGGCCACAGTTTGTG 960
CAGCGCTTCAACATTGCCAAAGGTGGCGAGTTCTACGGCG 1000
      1010      1020      1030      1040
.....
CCACCGAGGGTAATGCGAAGCATCATGAATCATGACAACAC 1040
GGTGGGGGCCATCGGCTTTTGTGTGGCGCATCCTGCCCAAG 1080
ATCTACCCAAATCTCGATCATTCGGGCGCATCCGGACACCG 1120
GAGAGCCCATTAGAGATAGGAATGGCCTATGCCAACTGTG 1160
CGCTCCCAACGAGGCCAGGCGTATTTCATCGGCAAGATCGTC 1200

```

Fig. 72A

1210	1220	1230	1240
AAAGGAAATCCTTCTCGCGAATTCTCGGATACGTCGATG	1240		
AAAAGGCCTCCGCGAAGAAGATTGTTAAGGATGTGTTCAA	1280		
GCATGGCGATATGGCTTTTCATCTCCGGAGATCTGCTGGTT	1320		
GCCGACGAGAAGGGTTATCTGTACTTCAAGGATCGCACCG	1360		
GTGACACCTTCCGCTGGAAGGGCGAGAATGTTTCCACCAG	1400		
1410	1420	1430	1440
CGAGGTGGAGGCGCAAGTCAGCAATGTGGCCGGTTACAAG	1440		
GATACCGTCGTTTACGGCGTAACCATTCGCGACACCGAGG	1480		
GAAGGGCGCGCATGGCCGCCATCTATGATCCGGAGCGAGA	1520		
ATTGGACCTCGACGCTCTTCGCCGCTAGCTTGGCCAAGGTG	1560		
CTGCCCGCGTACGCTCGTCCCCAGATCATTGCGATTGCTCA	1600		
1610	1620	1630	1640
CCAAGGTGGACCTGACTGGAACCTTTAAGCTGGCGCAAGGT	1640		
AGACCTGCAGAAGGAGGGCTACGATCCGAACGGGATCAAG	1680		
GACGCGCTGTACTACCAGACTTCCAAGGGTGGGTACGAGC	1720		
TGCTCACGCCCCAGGTTTACGACCCAGGTGCAGCGCAACGA	1760		
AATCCGCTTCTAAGAGCTGCAATAGAGTTGTGTCTGAACC	1800		
1810	1820	1830	1840
TTGCCCTTTTGCCCAATATGCTGTTAATTAGTTTGTAAAGC	1840		
TAAGTGTAGTAGAGGAAAATCGGGGGAAATCGGCAGCAAA	1880		
GATCATTACAGCCTAGGAGAGATGCATCCGAAGCACATTTT	1920		
CATGTCAACAATGCACCTTTTGTATATCGTAAGCATATATA	1960		
TATCGTATATCGTAAACGTAGTTGTATCTGCATTTGTGTA	2000		
2010	2020	2030	2040
GATGATAGCCCTCTATATGCAATTTCAATTGTTTTAGCGT	2040		
GCTAAAGAACTTGTAAATGCAATTTGAGCTATTGTTTA	2080		
GTCAGTTTTAGTGGCATTTACACTTCCATTCTCGTTGCGT	2120		
TTGCTTTTTGCGTGTACATATGAGAAAGCTCTGATGTTTTT	2160		
GTATCAAAATAAAGTTTTTTTCTTCAACACGGACCGTGA	2200		
2210	2220	2230	2240
AAAAAAAAAAAAAAAAAAAAA	2221		

Fig. 72B

```
      10      20      30      40
.....|.....|.....|.....|
ALWAYIKLLRYTKRHERLNYTVADVFERNYQAHFQKVAVV 40
SETGRWTFRGVNEHANKVANVLQAGGYKKGOVVALLLENR 80
AEYVATWLGLSKIGVITPLINTNLRGFSLLHSITVAHCSA 120
LIYGEDFLEAVTOVAKOLPANLTLFCFNNENNSETEKNI 160
PQAKNLNALLTTASYEKFNKTQVNHHCCLVYIYTSQTTGL 200
      210      220      230      240
.....|.....|.....|.....|
PKAAVISHSRYLFIAAGIHYTMGFQEEQIFYTPLPLYHTA 240
GGIMCMGQSVLFGSTVSIRKKFSASNYFADCAKYNATIGQ 280
YIGEMARYILATKPFSEYDCKHRVRLVFGNGLRPGIWFQFV 320
GRFNIAXVGEFYGATEGNANIMNHONTVGAIGFVSRILPK 360
IYFISIIIRADPCTGEPIDRNGLCQLCAPNEFGVFQKIV 400
      410      420      430      440
.....|.....|.....|.....|
KGNFSREFLGYYDEKASAKKIVKDVFKHGOMAFISGCLLV 440
ADEKGYLYFKORTGDTFRWKGENVSTSEVEAGVSNVAGYK 480
QTVVYGYTIFHTEGRAGMAAIYDFERELOLDVFAASLAKV 520
LPAYARPGIIRLLTKVCLTGTFKLRKVLOKEGYDFNAIK 560
DALYYCTEKGRYELLTPCYDQVCRNEIRF 590
```

Fig. 73

10 20 30 40  
.....|.....|.....|.....|  
AGTGTAGATACCACAGGAACGTTTAAAATCCAGAAGACCA 40  
GACTGCAAAGGGAAGGATACGATCCACGGCTCACAACCTGA 80  
CCAGATCTACTTCCTAAACTCCAGAGCAGGGCGTTACGAG 120  
CTTGTCAACGAGGAGCTGTACAATGCATTTGAACAAGGGC 160  
AGGATTTCCCTTT 173

Fig. 74

10 20 30 40  
.....  
SVDTTGTFKIQKTRLGREGYDFRLTTDQIYFLNSRAGRYE 40  
LVNEELYNAFEQGQOFF 57

Fig. 75

10 20 30 40  
ATGAAGCTGGAGGAGCTTGTGACAGTTATGCTTCTCACAG 40  
TGGCTGTCATTGCTCAGAATCTTCCGATTGGAGTAATATT 80  
GGCTGGAGTTCTTTATTTTATACATCACAGTGGTTCATGGA 120  
GATTTTCATTTATAGAAGTTATCTTACGTTGAATAGGGATT 160  
TAACAGGATTGGCTCTAATTATTGAAGTCAAAATCGACCT 200  
210 220 230 240  
ATGGTGGAGGTTGCATCAGAAATAAGGAATCCATGAACTG 240  
TTTTTGGATATTGTGAAAAAGAATCCAAATAAGCGGGGGA 280  
TGATTGACATCGAGACGAATACAAACAGAAACATAGGCAGA 320  
GTTCAATGCACATTGTAATAGATATGCCAATTATTTCCAG 360  
GGTCTTGGCTATCGATCCGGGAGACGTTGTCCGCTTGTACA 400  
410 420 430 440  
TGGAGAACTCGGTGAGTTTGTGGCCGCGTGGATGGGACT 440  
CGCAAAAATCGGAGTTGTAAACGGCTTGGATCAACTGGAAT 480  
TTGAAAAGAGAGCAACTTGTTCATTGTATCACTGGGAGCA 520  
AGACAAAGGCGATTATCACAAAGTGTAAACACTTCAGAAAT 560  
TATGCTTGGATGCTATCGATCAGAAAGCTGTTTGTATGTTGAG 600  
610 620 630 640  
GGAATTGAGGTTTACTCTGTGCGGAGAGCCCCAAGAAGATT 640  
CTGGATTCAAGAATCTCAAGAAGAAATTGGATGCTCAAAAT 680  
TACTACGGAAACCAAGACCCCTTGACATAGTAGATTTTAAA 720  
AGTATTCTTTGCTTCATCTATACAAAGTGGTACTACTGGAA 760  
TCCCAAAAAGCGCTGTCTATGAAGCACTTCAGATATTACTC 800  
810 820 830 840  
GATTGGCGTTGGAGCCGCAAAATCATTGCGGAATCCGCGCT 840  
TCTGATCGTATGTACGTCCTCGATGCCAATTTATCACACTG 880  
CAGCTGGAATTTCTGGAGTTGGGCAAGCTCTGTTGGGTGG 920  
ATCATGCTGTGTCATTAGAAAAAAATTCTCGGCTAGCAAC 960  
TTTTGGAGGGATTGTGTAAAGTATGATTGTACAGTTTCAC 1000  
1010 1020 1030 1040  
AATACATTGGAGAGATTTGTGCGGTACTTGTGGCTCAGCC 1040  
AGTTGTGGAAAGAGGAATCCAGGCATAGAATGAGATTGTTG 1080  
GTTGGAAACGGACTCCGTGCTGAAATCTGGCAACCAATTG 1120  
TAGATCGATTCCGTGTGTCAGAAATGGAGAACTTTATGGTTT 1160  
AACTGAAGGAACATTCATCTCTCGTGAACATTGACGGACAT 1200

Fig. 7A

```
      1210      1220      1230      1240
.....|.....|.....|.....|
GTCGGAGCTTGCGGATTCTTGCCAATATCCCCATTAACAA 1240
AGAAAAATGCATCCGGTTCGATTAATTAAGGTTGATGATGT 1280
CACTGGAGAAGCAATCCGAACCTCCGATGGACTTTECATT 1320
GCATGTAATCCAGGAGAGTCTGGAGCAATGGTGTCCACGA 1360
TCAGAAAAAATAATCCATTATTGCAATTTCGAGGGATATCT 1400
      1410      1420      1430      1440
.....|.....|.....|.....|
GAATAAGAAGGAAAACGAATAAAAAAGATTATCAGAGATGTC 1440
TTCCGAAAGGGAGATAGTTGCTTTTGGACTGGAGATCTTC 1480
TTCATTGGGATCGTCTTGTTTATGTATATTTCAAGGATCG 1520
TACTGGAGATACTTTCCGTTGGAAGGGAGAGAATGTGTCC 1560
ACTACTGAAGTCGAGGCAATTCTTCATCCAATTACTGGAT 1600
      1610      1620      1630      1640
.....|.....|.....|.....|
TGTCGTGATGCAACTGTTTATGGTGTAGAGGTTCTCTCAAAG 1640
AGAGGGGAAGAGTTGGAATGGCGTCAGTTGTTCCAGTTGTA 1680
TCGCATGAGGAAGATGAAACTCAATTTGTTTCATAGAGTTG 1720
GAGCAAGACTTGCCTCTTCGCTTACCAGCTACGCGATTCC 1760
TCAGTTTATGCGAATTTGTCAGGATGTTGAGAAAAACAGGT 1800
      1810      1820      1830      1840
.....|.....|.....|.....|
ACATTCAAACCTTGTGAAGACGAATCTACAACGATTAGGTA 1840
TCATGGATGCTCCTTCAGATTCAATTTACATCTACAATTG 1880
TGAAAAATCGCAATTTTGTGCGGTTGACAAATGATTTGAGG 1920
TGCAAGGTCTCACTGGGAAGTTATCCATTTTAA 1953
```

Fig. 74B

```
      10      20      30      40
.....
MKLEELVTVMLLTVAVIAGNLPIGVILAGVLILYITVYHG 40
DFIYRSYLTNLNROLTGLALIEVKIDLWWRLHCKNGHEH 60
FLDIYKKNPKNKPFAMIDLETNTTETTYAEFNAHCNRYANYFO 120
GLGYRSGDVVALYMENSVEFVAAWMGLAKIGVVTAWINSN 160
LKREQLVHCITASKTKAIIISVTLQNIMLQAIQCKLFQVE 200
      210      220      230      240
.....
GIEVYSVGEFKKNSGFKNLKKKLQAGITTEPKTLDIVDFX 240
SILCFIYTSGTTGMPKAAVKNKHFRYYSAVGAAXSFGIRF 260
SDRMVYSMPHYHTAAGILGVGCALLGGSSCVIRKKFSASN 320
FWROCVKYDCTVSGYIGEICRYLLAQPVVEEESRHRMRL 360
VGNGLRAEIWCPFVORFRVRIGELYGSTEGTSLLVNIQGH 400
      410      420      430      440
.....
VGACGFLPISPLTKKMHFVRLIKVDDVTGEAIRTSGLCI 440
ACNFGESGAMYSTIRKNNPLLQFEGYLNKKETNKKIIRCV 480
FAKGSQCFLTGOLLHWQRLGYVYFKDRTGDTFRWKGENVS 520
TTEVEAILHPITGLSDATVYGVEVPQREGRYGNASVVRVV 560
SHEEDETCFVHRVGARLASSLTSYAIPQFMRICDDVEKTS 600
      610      620      630      640
.....
TFKLVKTNLGRLGIMDAPSDSIYLYNSENRNPFVFFQDLR 640
CKVSLGSYFF 650
```

Fig. 77



10 20 30 40  
ATGAGGGAAATGCCGGACAGTCCCAAGTTTGCGTTAGTCA 40  
CGTTTGTGTGTATGCAGTGGTTTTGTACAATGTCAACAG 80  
CGTTTTCTGGAAATTTGTATTCATCGGATATGTTGTATTT 120  
AGGCTGCTTCGCACTGATTTTGGAAAGAAGAGCACTTGCCA 160  
CGTTACCTAGAGATTTTGCGGGACTGAAGCTCTTAATATC 200  
210 220 230 240  
GGTTAAGTCGACAATTTCGTGGCTTGTTCAGAAAGATCCG 240  
CCAATTCATGAAATCTTTTTGAATCAGGTGAAACAGCATC 280  
CAAAACAAAGTGGCGATTATTGAAATTGAAAGTGGTAGGCA 320  
GTTGACGTATCAAGAATTGAATGCGTTAGCTAATCAGTAT 360  
GCTAACCTTTACGTGAGTGAAGGTTACAAAATGGGCGACG 400  
410 420 430 440  
TTGTGCGCTTTGTTTATGGAAAAATAGCATCGACTTCTTTGC 440  
AATTTGGCTGGGACTTTCCAAGATTGGAGTCGTGTGGGCG 480  
TTCATCAACTCAAACTTGAAGTTGGAGCCATTGGCACATT 520  
CGATTAATGTTTGAAGTGCAAAATCATGCAATTACCAATAT 560  
CAATCTGTTGCGGATGTTCAAAAGCCGCTCGTGAAGAAGAT 600  
610 620 630 640  
CTGATCAGTGACGAGATCCACGTGTTTTCTGGGTGGAACTC 640  
AGGTTGATGGACGTGATAGAAGTCTTCCAGCAAGATCTCCA 680  
TCTTTTCTCTGAGGATGAACCTCCAGTTATAGACGGACTC 720  
AATTTTGAAGCGTTCTGTGTTATATTTACACTTCCGGTA 760  
CTACCGGAAATCCAAAGCCAGCCGTCATTAAACACTTCCG 800  
810 820 830 840  
TTACTTCTGGATTGCCATGGGAGCAGGAAAAGCATTTTGA 840  
ATTAATAAGTCAGACGTTGTGTACATTACGATGCCAATGT 880  
ATCACTCTGCCGCCGGTATCATGGGTATTGGATCATTAAAT 920  
TGCATTGGGTGGACCGCTGTTATTAGGAAAAAGTTTTCG 960  
GCAAGCAACTCTTGGAAAGATTGGGTCAAGTACAACGTCA 1000  
1010 1020 1030 1040  
CAGCGACACAGTACATTGGAGAAATCTGCAGGTATCTTCT 1040  
GGCAGCGAATCCATGTCTTGAAGAGAAACAACACAAAGTG 1080  
CGATTGATGTGGGGAATGGTTTGAAGAGGACAAATTTGGA 1120  
AAGAGTTTGTAGGAAGATTTGGAATTAAGAAAAATTTGAGA 1160  
GTTGTACGGCTCAACAGAAAGGAAACTCCAATATTGTTAAC 1200

Fig. 78A

1210	1220	1230	1240
GTGGATAACCATGTTGGAGCTTGTGGATTTCATGCCAATTT	1240		
ATCCCCATATTGGATCCCTCTACCCAGTTCCGACTTATTAA	1250		
GGTTGATAGAGCCACTGGAGAGCTTGAACGTGATAAGAAC	1320		
GGACTCTGTGTGCCGTGTGTGCCTGGTGAAACTGGGGAAA	1360		
TGGTTGGCGTTATCAAGGAGAAAGATATTCTTCTAAAGTT	1400		
1410	1420	1430	1440
CGAAGGATATGTCAECGAAGGGGATACTGCAAAGAAAATC	1440		
TACAGAGATGTGTTCAAGCATGGAGATAAGGTGTTTGCAA	1480		
GTEGAGATATTCTTCATTGGGATGATCTTGGATACTTGTA	1520		
CTTTGTGGACCGTTGTGGAGACACTTTCCGTTGGAAAGGG	1560		
GAGAACGTGTCAACTACTGAAGTTGAGGGAAATCTTCAGC	1600		
1610	1620	1630	1640
CTGTGATGGATGTGGAAGATGCAACTGTTTATGGAGTCAC	1640		
TGTCCGTAAAATGGAGGGGGCGTGCCGGAATGGCTGGTATT	1680		
GTCGTCAAGGATGGAACGGATGTTGAGAAATTCATCGCCG	1720		
ATATTACTTCTCGACTGACCGAAAAATCTGGCGTCTTACGC	1760		
AATCCCTGTTTTTCATTCCGCTGTGCAAGGAAGTTGATCGA	1800		
1810	1820	1830	1840
ACCGGAACCTTCAAACTCAAGAAGACTGATCTTCAAAAAC	1840		
AAGGTTACGACCTGGTTGCTTGTAAGGAGACCCAATTTA	1880		
CTACTGGTCAGCTGCAGAAAAATCCTACAAACCACTGACT	1920		
GACAAAATGCAACAGGATATTGACACTGGTGTTTATGATC	1960		
GCATTTAA	1966		

Fig. 78B

10	20	30	40
MREMFDSPKFALVTFVYYAVVLYNVNSVFWKVFVFIGYVVF	40		
RLLRTDFGRRALATLPROFAGLKLLISVKSTIRGLFKKOR	80		
PIHEIFLNQVKQHFNKVAITEIESGRGLTYOELNALANGY	120		
ANLYVSEGYKMGQVVALFMENSIOFFAIWLGLSKIGVYSA	160		
FINSNLKLEPLAHSINYSKCKSCITNINLLFMFKAAREKN	200		
210	220	230	240
LISGEIHVFLAGTQVQGRHRSLOQDLHLFSEDEFPYIDGL	240		
NFRSVLCYIYTSGTTGNPKPAVVKHFRYPWIMGAGKAFG	280		
INKSCVYVITMPMYHSAAGIMGIGSLIAFGSTAVIRKKFS	320		
ASNFWKQCVKYNVTATQYIGEICRYLLAANFOPEEKQHN	360		
RLMWGNGLRGGIWKEFVGRFGIKKIGELYGSTEGNSNIVN	400		
410	420	430	440
VONHVGACGFMPITYPHIGSLYPVRLIKVORATGELERCKN	440		
GLCVFQVPGSTGEMVGVKEKQILLKFEQYVSEGGTAKKI	480		
YRCVFKHGCKVFASGOILHWQDLGYLYFVOROGOTFRWK	520		
ENVSTTEVEGILQPYMVEEDATVYGVTVGKMEGRAGMAG	560		
VVKCGTQVEKFIACITSRLTENLASVAIPVFIRLCKEVD	600		
610	620	630	640
TGTFXLLKKTCLQKGGYGLVACKGDFIYYWSAAEKSYKFLT	640		
DKMCCGIDTGVYDR	655		

Fig. 79

10 20 30 40  
ATGGCGTGATGCATCAGGCTCAGCTATACAATGATCTAG 40  
AGGAATTGCTAACTGGTCCATCAGTACCCATCGTTGCTGG 80  
AGCTGCTGGAGCTGCAGCTCTCACTGCCCTACATTAACGCC 120  
AAATACCACATAGCCCATGATCTCAAGACCCCTCGGTGGTG 160  
GATTGACACAATCGTCCGAAGCGATTGATTTTATAAACCG 200  
210 220 230 240  
CCGCGTGGCACAAGAGCGCGTCTCAGGCACACATCTTC 240  
CAGGAGCAGGTCCAAAAACAATCAAATCATCCCTTTCTTA 280  
TCTTTGAGGGCAAGACATGGTCTTACAAGGAGTTCTCTGA 320  
GGCATACAGGAGGGTGGCGAACTGGCTGATTGATGAGCTG 360  
GACGTACAAGTAGGGGAGATGGTGGCAATTGATGGGGGAA 400  
410 420 430 440  
ATAGTGCAGAGCACCTGATGCTTTTGGCTTGGACTTGAAG 440  
AATCGGTGCGGGTACGAGTTTTTTGAAGTGGAACTGACA 480  
GGGGCAGGGTTAATTCATTGCAATAAAGCTATGGGAATGTC 520  
GATTCGTTATCGCAGACATCGATATTTAAAGCGAATCATTGA 560  
ACCGTGGCGTGGCGAACTGGAGGAGACGGGCATCAACATT 600  
610 620 630 640  
CACTACTATGACCCATCCTTTTCATCTCATCGGTACCGAATA 640  
ACACGGCAATTCGGGACAGCGCGCACTGAGAAATTTGAATT 680  
AGATTGAGTACGAGGACTGATATACACATCTGGAAACCACT 720  
GGTCTACCTAAAGGGCTGTTTATAAGCACTGGCGCGGAGCG 760  
TTAGGACTGACTGGTGGATTTCAAAGTATCTAAATCTCAA 800  
810 820 830 840  
GCCACGGATCGAATGTATACATGTATGCGGCTCTACCAT 840  
GCCGCTGCACACAGCCTCTGTACAGCATCAGTTATTCATG 880  
GTGGAGGTACCGTGGTATTGAGCAGGAAATTCACACAAA 920  
GAAGTTCTGGCCTGAAGTTGTGGCTTCGGAAAGCAATATC 960  
ATTCAGTACGTTGGTGAATTAGGTCGATATCTCTTGAATG 1000  
1010 1020 1030 1040  
GTCCAAAGAGTCTTTACGACAGGGCCCCATAAAGTCCAGAT 1040  
GGCGTGGGGCAATGGCATGGCTCCAGACGTGTGGGAAGCG 1080  
TTTCGTGAACGCTTCAACATAACCAATTATTCATGAGCTCT 1120  
ATGCCGCAACCGATGGGCTCGGGTCAATGACCAATCGTAA 1160  
CGCGGGCCCTTTTACAGCAAACTGTATTGGCTGGGAGGG 1200

Fig. 20A

```
      1210      1220      1230      1240
.....|.....|.....|.....|
CTGATCTGGCACTGGAAATTTCGAAATCAGGAAGTGCTGG 1240
TCAAGATGGATCTCGATACTGATGAGATCATGAGAGATCG 1280
CAATGGGTTTGGGATACGATGCGCTGTCAATGAACCTGGA 1320
CAGATGCTTTTTCGGCTGACACCCGAAACTCTGGCTGGTG 1360
CACCAAGCTACTACAACAACGAAACGGCCACACAGAGCAG 1400
      1410      1420      1430      1440
.....|.....|.....|.....|
GCGGATTACAGATGTGTTTCAAAAGGGTGACCTGTGGTTC 1440
AAGTCCGGTGACATGCTACGGCAAGACGCCGAAGGCCGCG 1480
TCTACTTTGTGATCGACTAGGCGATACGTTCCGCTGGAA 1520
ATCCGAAAACGTTTCTACCAATGAAGTCGCGGACGTGATG 1560
GGCACATTTCTCTCAGATTGCTGAAACGAATGTATACGGTG 1600
      1610      1620      1630      1640
.....|.....|.....|.....|
TCCTTGTCGCGGGTAACGATGGTCCAGTGCGCAGCCTCAA 1640
TTGTCATGGCAGACGGCGTGACAGAGTCGACATTGCTTC 1680
GCTGCCCTTGCAAAGCACGCCCGAGATCGGTTACCGGGTT 1720
ATGCTGTACCACTGTTTCTGAGGGTAACTCCAGCACTTGA 1760
ATATACGGGCACATTAAGATTTCAGAAAGGACGCCTCAAG 1800
      1810      1820      1830      1840
.....|.....|.....|.....|
CAGGAAGGTATAGACCCAGATAAGATTTCCGGCGAAGATA 1840
AGTTATACTGGCTGCCGCCCTGGTAGCGATATATATTTACC 1880
ATTGGAAGAAGATGGAGTGGCAGGGAATTGTAGATAAGCGT 1920
ATACGGCTGTGA 1932
```

Fig. 80B

10	20	30	40
MACHHCAQLYNQLEELLTGFSVP	I VAGAAGAAAALTAY	INA	40
KYHIAHDLKTLGGGLTQSSEA	IDF INRRVAQKRVLT	HHIF	80
QEQVQKCSNHPFLIFEGKTWSYKEFSEAYTRVANWL	IGEL		120
OYQVGENVAIDGGNSAEHLMLWLALDAIGAATSFLNWNLT			160
GAGLIHCILKLCGRFVIAOIOIKANIEPCRGEEETGIN			200
210	220	230	240
HYYPFSFISSLFNNTPIPOSRTENIELOSVRGLIYTSSTT			240
GLPKGVFISTGRELRTOWSISKYLNKPTORMYTOMFLYH			280
AAAHSLCTASVIEGGGTIVLSRKFSKKFWPEVVASEAN			320
IQYVGELGRYLLNGPKSPYDRAHKVQMAWGNGMRPOVWEA			360
FRERFNPIIHELYAATDGLGSMTNRNAGFFTANCIALRG			400
410	420	430	440
LIWHWKFRNGEVLVKMCLDTGEIMRDRNGFAIRCAYNEFG			440
QMLFRLTPETLAGAPSYNNETATOSRRITOVFQKGLWF			480
KSGGMLRGDAEGRVYFVORLGOTFRWKSENVSTNEVADVM			520
GTFFQIAETNVYGVLPQNDGRVRSLNCHGRRRQRYD			560
AAAKHARORLFGYAVPLFLRVTPALEYTGTLKIQKGRUK			600
610	620	630	640
CEGIGPKISGECKLYWLPFGSDIYLPFGKMEWIGIVOKR			640
IRL			640

Fig. 81

```
      10      20      30      40
.....|.....|.....|.....|
CTTTACCATTTCATCAGCTTCATTCTGCATTTTGTAGCTTGA 40
CGGCAGCCCGGCTCTACGCTGATCATCGGCCGCAAGTTCTC 80
CGCGAGAACTTCATAAAGGAAGCGCGCGAGAACEACGCC 120
ACGGTCATCCAGTACGTGGGTGAGACCTTTCGATATCTGC 160
TCGCCACCCCCCGGTGAAACCGATCCAGTTACTGGCGAAGA 200
      210      220      230      240
.....|.....|.....|.....|
CCTGGACAAAAAGCACAAATATTCGAGCAGTATACGGCAAC 240
GGGCTACGGCCGGATATCTGGAACCGCTTCAAGGAGCGCT 280
TCAACGTGCCGACGGTTGCCGAATTTTATGCTGCAACCGA 320
GAGCCCAGGCGGAACATGGAACATTTCAACAAATGACTTC 360
ACTGCCCGAGCCATTGGGCGACACTGGCGTGCTTAGTGGAT 400
      410      420      430      440
.....|.....|.....|.....|
GGCTTCTTGGACGCGCGCTTACTATTGTGAGGTGGACCA 440
GGAATCACAGGAACCATGGCGCGATCCCCAAACCGGGTTC 480
TGCAAGCCGGTCCCGCGAGGCGAAGCAGGCGAGCTCCTGT 520
ATGCCATTGATCCGGCCGACCCGGGCGAGACCTTCCAGGG 560
CTACTACCGCAACTCCTTTAGAGCACACTGGCGGCGG 600
```

Fig. 82

10 20 30 40  
.....  
LYHSSASFCIFSLTAAGSTLIIGRKFSARNFIKEARENCA 40  
TVIQYVGETLRYLLATPGETDPVTGEDLOKKHNIRAVYGN 80  
GLRFOIWNRFKERFNVFTVAEFYAATESPGGTWNYSTNOF 120  
TAGAIGHTGVLSGWLLGRGLTIVEVDQESQEPWROPQTGF 160  
CKFVFRGEAGELLYAIOPAOPGETFQGYRNSFRAHWRF 199

Fig. 83



```
      10      20      30      40
.....|.....|.....|.....|
GCAAAGGCCGACGCGTGECTGCGGACGGGTAACGTGATCA 40
GGGCGGACAACGAAGGGCGEACTCTTCTTCCACGACCGGAT 80
CGGAGACACGTTCCGATGGAAGGGAGAGACNCTCAGCACA 120
CAAGAGGTCAGTTTGGTGCTCGGACGACACGACTCAATCA 160
AGGAGGCCAACGTGTACGGCGTGACGGTGCCGAACCACGA 200
      210      220      230      240
.....|.....|.....|.....|
CGGGCGGGCCGGCTGCGCTGCGCTCACGCTATCAGACGCT 240
CTGGCGACTGAAAAGAAGCTGGGCGATGAGCTGCTAAAGG 280
GATTGGCTACTCACTCGTCGACTTCGCTTCCCAAGTTTGC 320
GGTGCCGCAGTTCCCTACGGGTGGTGCGCGGCGAGATGCAG 360
TCAACGGGCACCAACAAGCAACAGAAACACGACCTGAGGG 400
      410      420      430      440
.....|.....|.....|.....|
TGCAGGGTGTAGAGCCGGGCAAGGTGGGCGTAGACGAGGT 440
GTACTGGTTGCGGGGAGGGACATATGTACCATTGGAACA 480
GAGGATTGGGATGGGTTGAAGAAGGGCTCTGTGAAGTTGT 520
GA 522
```

Fig. 84

10 20 30 40  
.....  
AKADAWLRTGNVIRADNEGRLEFFHDIRIGOTFRWKGETYST 40  
QEVSLVLGRHDSIKEANYVGVTVFNHQGRAGCAALTLSOA 80  
LATEKKLGDELLKGLATHSSTSLFKFAVPGFLRYVRGEMQ 120  
STGTNKQCKHQLRVQGVFPGKVGVDDEVYWLRGGETYYFFGT 160  
EDWDGLKKGLVKL 173

Fig. 85

```

      10      20      30      40
.....
ATGTCGCCCATACAGGTTGTTGTCTTTGCCCTTGTCAAGGA 40
TTTTCTGCTATTATTTCAGACTTATCAAGCTAATTATAAC 80
CCCTATCCAGAAATCACTGGGTTATCTATTTGGTAATTAT 120
TTTGATGAATTAGACCGTAAATATAGATACAAGGAGGATT 160
GGTATATTATTCTTACTTTTTGAAAAGCGTGTTTTGTTA 200
      210      220      230      240
.....
TATCATTGATGTGAGAAGACATAGGTTTCAAACTGGTAC 240
TTATTTATTAAACAGGTCCAACAAAATGGTGACCATTTAG 280
CGATTAGTTACACCCGTCCCATGGCCGAAAAGGGAAGATT 320
TCAACTCGAAACCTTTACGTATATTGAAACTTATAACATA 360
GTGTTGAGATTGTCTCATATTTTGCATTTTGATTATAAAG 400
      410      420      430      440
.....
TTCAGGCGCGTGACTACGTGGCAATCGATTGTACTAATAA 440
ACCTCTTTTGGTATTTTTATGGCTTTCTTTGTGGAAACATT 480
GGGGCTATTCCAGCTTTTTTAAACTATAATACTAAAGGCA 520
CTCCGCTGGTTCACCTCCCTAAAGATTTTCCAATATTACGCA 560
GGTATTTATTGACCCGTGATGCCAGTAATCCGATCAGAGAA 600
      610      620      630      640
.....
TCGGAAGAAGAAATCAAAAACGGCACTTCTGTGTTAAAT 640
TAAACTATCTTGAGAACAAGACTTAATGCATGAACTTTT 680
AAATTCGCAATCAGCGGAATTCTTACAACAAGACAAAGTT 720
AGGACACCACTAGGCTTGACCGATTTTAAACCCCTCTATGT 760
TAATTTATACATCTGGAAACCACTGGTTTGGCTAAATCCGC 800
      810      820      830      840
.....
TATTATGTCTTGGAGAAAAATCCTCCGTAGGTTGTCAAATT 840
TTTGGTCATGTTTACATATGACTAATGAAAGCACTGTGT 880
TCACAGCCATGCCATTGTTCCATTCAACTGCTGCCTTATT 920
AGGTGCGTGCGCCATTCTATCTCAAGGTGGTTGCCTTGCG 960
TTATCGCATAAATTTTCTGCCAGTACATTTTGGAAAGCAAG 1000
      1010      1020      1030      1040
.....
TTTATTTAAACAGGAGCCACGCACATCCAATATGTGGGAGA 1040
AGTCGTGTAGATACCTGTTACATACGCCAATTTCTAAGTAT 1080
GAAAAGATGCATAAGGTGAAGGTTTGCTTATGGTAACGGGC 1120
TGAGACCTEACATCTGGCAGGACTTCAGGAAGAGGTTCAA 1160
CATAGAAGTTATTGGTGAATTCATGCGCGCAACTGAAGCT 1200

```

Fig. 86A

```
      1210      1220      1230      1240
.....
CCTTTTGCTACAACCTTCCAGAAAGGTGACTTTGGAA 1240
TTGGCGCATGTAGGAACTATGGTACTATAATTCAATGTT 1280
TTTGTCAATCCAAACAAACATTGGTAAGGATGGACCCAAAT 1320
GACGATTCCGTTATATATAGAAATTCCAAGGGTTTCTGCG 1360
AAGTGGCCCCCTGTTGGCGAACCAGGAGAAATGTTAATCAG 1400
      1410      1420      1430      1440
.....
AATCTTTTTCCCTAAAAAACCAGAAACATCTTTTCAAGGT 1440
TATCTTTGTAATGCCAAGGAAACAAAGTCCAAAGTTGTGA 1480
GGGATGTCTTCAGACGTGGCGATGCTTGGTATAGATGTGG 1520
AGATTTATTAAAAAGCGGACGAATATGGATTATGGTATTTT 1560
CTTGATAGAATGGGTGATACTTTTCAGATGGAAATCTGAAA 1600
      1610      1620      1630      1640
.....
ATGTTTCCACTACTGAAGTAGAAGATCAGTTGACGGCCAG 1640
TAACAAAGAAACAATATGCACAAGTTCTAGTTGTTGGTATT 1680
AAAGTACCTAAATATGAAGGTAGAGCTGGTTTTTGCAGTTA 1720
TTAAACTAACTGACAACCTCTCTTGACATCACTGCAAGAGAC 1760
CAAAATTATTAAATGATTCCCTTGAGCCGGTTAAATCTACCG 1800
      1810      1820      1830      1840
.....
TCTTATGCTATGCCCTATTTTGTAAATTTGTTGATGAAA 1840
TTAAATGACAGATAACCTCATAAAATTTTGA 1872
```

Fig. 86B

10	20	30	40
.....			
MSPIQVVVFALSRIFLLLFRLIKLIITPICKSLGGLFGNY	40		
FDLDRKYRYKEDWYIIPYFLKSVFCYIDVRRHRFQNWY	60		
LFIKQVQNGGDLAISYTRFMASKGEFCLSTPTTYIETYN	120		
VLRLSHILHFDYNYGAGDYVAIDCTNKPLFVFLWLSLWN	180		
GATPAFLNYNTKGTPLVHSLKISNITQVFIDPDASNPIRE	200		
210	220	230	240
.....			
SEEEIKNALFDVKLNYLEEEOOLMHSELLNSQSPFLOQDNV	240		
RTPLGLTQFKPSMLIYTSGLTGLPKSAIMSWRKSSVGGCV	280		
FGHYLHMTNESTVFTAMPLPHSTAALLGACATLSHGGLA	320		
LSHKFSASTFWKGVYLTGATHIQYVGEVCRYLLHTPIISKY	360		
EKMHKVKVAYGNGLRFOIWOOFKRFNIEVIGEPYAATEA	400		
410	420	430	440
.....			
FFATTTTFQKQFQIGACRNYGTICWFLSFQDTLVXQDPN	440		
QCSVIYRNSKGFCEVAFVGEPEMLMRIFPKKXRETGFQ	480		
YLGNAKETKSKVYRQVFRFGDAWYRCGOLLKAGEYSLWYF	520		
LDRNGOTFRWKSENVSTTEVEOCLTAENKEQYACVLYVQI	560		
KVPKYEGRAQFAVIKLTQNSLQITAKTKLLNGSLSRNLQ	600		
610	620	630	640
.....			
SYAMPLFVKFVDEIKMTDNLIK	623		

Fig. 87

10 20 30 40  
GTGTCGGATTACTACGGCGGGCGCACACACAACGGTCAGGC 40  
TGATCGACCTGGCAACTCGGATGCCGCGAGTGTGGCGGA 80  
CACGCCGGTGAATTGTGCGTGGGGCAATGACCGGGCTGCTG 120  
GCCCCGGCGAATTCCAAGCGTCGATCGGCACGGTGTTC 160  
AGGACCGGGCCGCTGCTACGGTGACCGAGTCTTCTGAA 200  
210 220 230 240  
ATTCGGCGATCAGCAGCTGACCTACCGCGACGCTAACGCC 240  
ACCGCCAAACCGGTACGCCGCGGTGTGGCCGCGCGCGCG 280  
TCGGCCCCGGGACGTCGTGGCATCATGTTGCGTAACTC 320  
ACCCAGGCACAGTCTTGGCGATGCTGGCCACGGTCAAGTGC 360  
GGCGCTATCGCCGGCATGCTCAACTACCACGAGCGCGCG 400  
410 420 430 440  
AGGTGTTGGCGCACAGCCTGGGTCTGCTGGACGGCGAAGGT 440  
ACTGATCGCAGAGTCCGACTTGGTCAGCGCGCTCGCGGA 480  
TGCGCGCGCTCGCGCGCGCGCGGTAGCGGGCGACGTCCTGA 520  
CGGTGAGGACGTGGAGCGATTCCGCACAAAGCGCGCGCG 560  
CACCAACCGCGCTCGCGCTCGCGCGTGCAGCGCAAAACAC 600  
610 620 630 640  
ACCGCGTTCTACATCTTTCACCTCGGGCACCAACGGATTTC 640  
CCAAGGCCAGTGTGATGACGCAATCATCGGTGGCTGCGGG 680  
GCTGGCGGTCTTTCGAGGGATGGGGCTGGGGTGAAGGGT 720  
TGGACACGGCTGTACAGCTGGCTGGCGCTGTACCAACAA 760  
ACCGGTTAACGGTCCGCGTGTGCTGGGTGAACAAATCTGG 800  
810 820 830 840  
GGCGACCTTGGCGCTGGGTAAAGTCGTTTTTCGGCGTGGCG 840  
TTCTGGGATGAGGTGATTGCCAATCGGGCGACGGCGTTGG 880  
TCTACATCGGCGAAATCTGCCGTTATCTGCTCAACAGCG 920  
GGCGAAGCGGACCGAACCCTGGCGACCGAGGTGGGGTGAATC 960  
TGGGTAAACGGGTGGCGCGGGAGATCTGGGATGAGTTCA 1000  
1010 1020 1030 1040  
CCACCCGCTTCGGGGTGGCGCGGGTGTGGGAGTTCTACGC 1040  
CGCCAGCGAAGGCAACTCGGCCCTTTATCAACATCTTCAAC 1080  
GTGCCCCAGGACCGCGGGGTATCGCGGATGGCGCTTGGCT 1120  
TTGTGGAATACGACCTGGACACCGCGGATCGGCTGGGGGA 1160  
TGGAGCGGGCGAGTGGCTCGGGTACCCGACGGTGAACCC 1200

Fig. 88A

```
      1210      1220      1230      1240
.....
GGCCTGTTGCTTAGCCCGGTC AACC GGCTGCAGCCGTTCC 1240
ACGGCTACACCGACCCGGTTGCCAGCGAAAAGAGTTGCT 1280
GCGCAACGCTTTTCGAGATGGCGACTGTTGGTTCAACACC 1320
GGTGACGTGATGAGCCCGCAGGGCATGGGCCATGCCGCT 1360
TCGTGATCGGCTGGGCGACACCTTCCGCTGGAAGGGCGA 1400

      1410      1420      1430      1440
.....
GAATGTCGCCACCACTCAAGTCGAAGCGGCACTGGCCTCC 1440
GACCAGACCGTCGAGGAGTGACGGTCTACGGCGTCCAGA 1480
TTCCGCGCACCGCGCGGGCGCGCGGAATGGCCGCGATCAC 1520
ACTGCGCGCTGGCGCGGAATTGACGCGCCAGGCGCTGGCC 1560
CGAACGGTTTACGGTCACTTGCCCCGCTATGCACTTCCGC 1600

      1610      1620      1630      1640
.....
TCTTTGTTCCGGTAGTGGGGTCGCTGGCGCACACCACGAC 1640
GTTCAAGAGTCGCAAGGTGGAGTTGCGCAACCAGGCCTAT 1680
GGCGCCGACATCGAGGATCCGCTGTACGTACTGGCCGCGC 1720
CGGACGAAGGATATGTGCCGTACTACGCCGAATACCCGTA 1760
GGAGGTTTCCGCTCGGAAGGCGACCGCAGGGCTAG 1794
```

Fig. 88B

```
      10      20      30      40
.....
MSDYGGGAHTTVRLIDLATRMFRVLAOTFVIVRCAMTGLL 40
ARPNKASIGTVFOORAARYGDRVFLKFGGGLTYROANA 80
TANRYAAVLAARGVGFGGVVGIMLRNRPSTVLANLATVKG 120
GATAGMLNYHGRGSEVLANSGLLDKVLIAESGLVSAYAE 160
CGASRGVAGGVLTVEOVERFATTAPATNPFASASAVCAKQ 200
      210      220      230      240
.....
TAFYIFTSGTTGFPKASVMTHHRWLALAVFGGGLRLKQ 240
SDTLYSCLPLYHNNALTVAVSSVINSGATLALCKSPSASR 280
FWGEVIANRATAFVYIGETCRYLLNCPAKFTDRAHCVRYI 320
CGNQLRPEIWDSEFTTRFGVARVCEFYAASEGNSAFININ 360
VPRTAGVSPMLAFVEYDLDTGDFLRDASGRVRRYPGGEP 400
      410      420      430      440
.....
GLLLSRVNRLOPFDGYTDFVASEKKLVRNAFRGGGOWENT 440
GDVMSPOEMCHAAEFVORLGSTFRWKGENYATTQVEAALAS 480
DCTVEECTVYGVQIFRTGGGRAGMAAITLRAGAEFDGQALA 520
RTVYGHLPQYALPLFVRVVGSLAHTTTFKSRKYELRNQAY 560
GADIEDFLYVLAQPDGCVVFPYAYEYFEEVSLGRFQG 596
```

Fig. 89



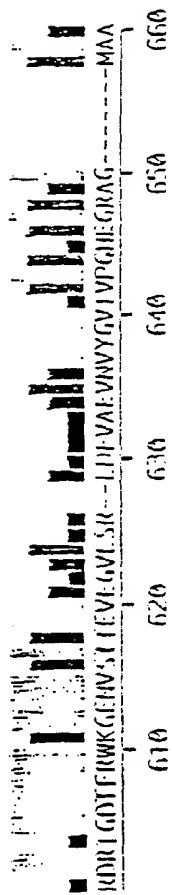
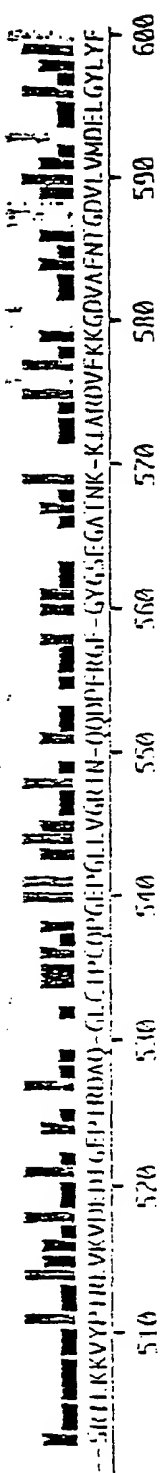
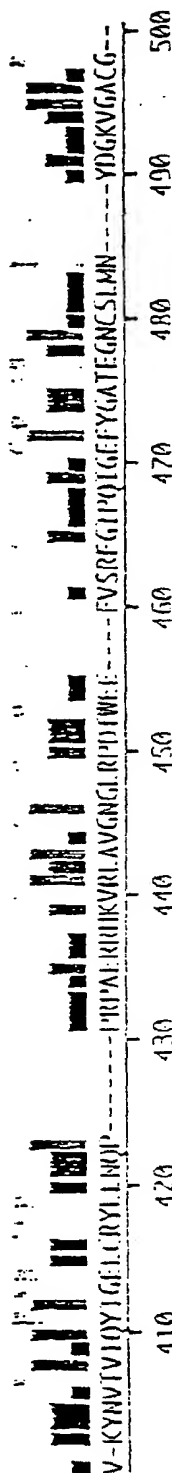
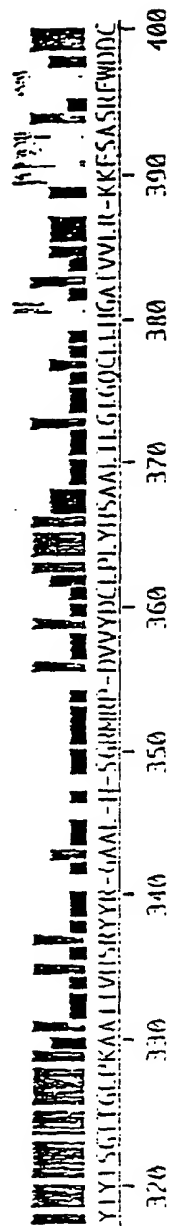


Figure 90

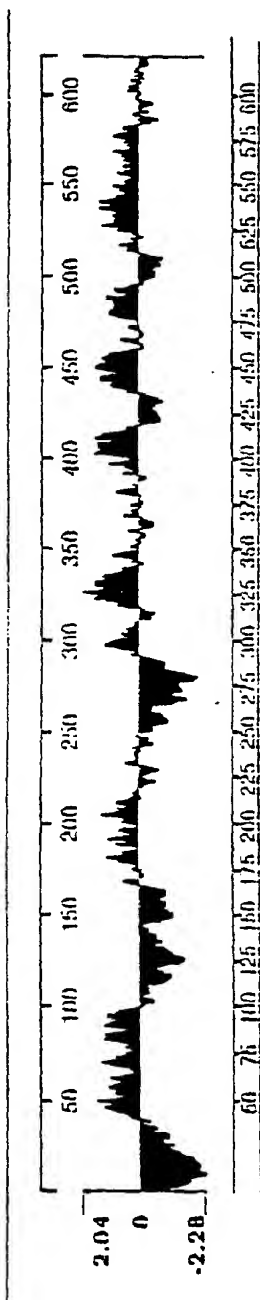


Figure 91

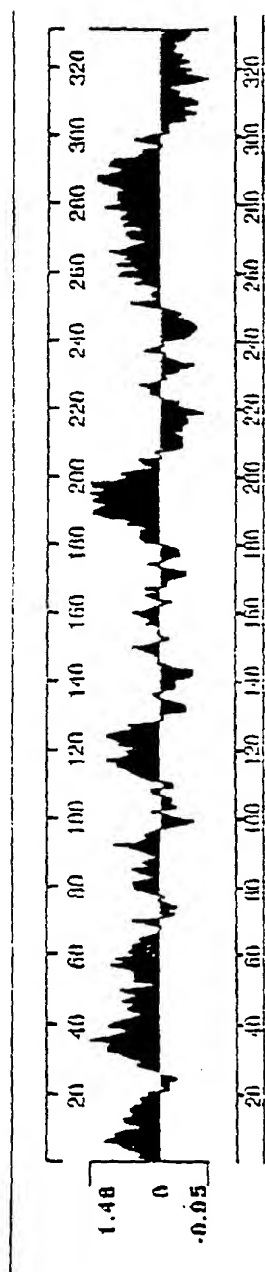


Figure 92

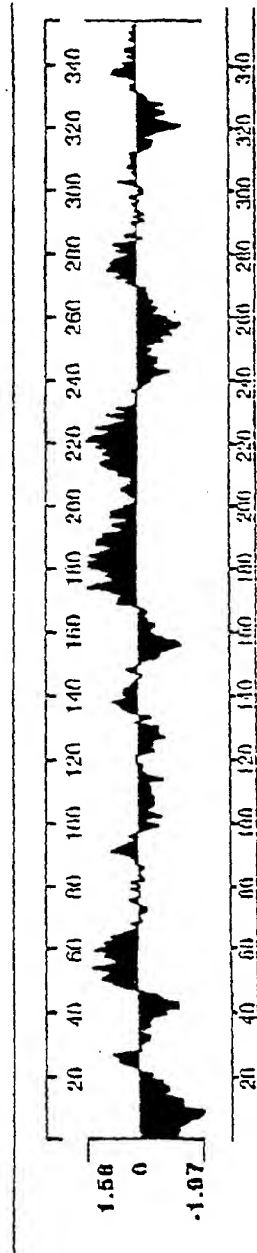


Figure 93

1892223

1      GGA GGC AGC GGT GGG GGG AGC GGC GGC GGC GGC AAC GGC AGC GGT GGC ACA GAG GAA  
1      M F A S G W H Q T V P I E E

61      GAG GGC GGC AGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC  
61      A G S H A A L L L L P L L L L L P L L L L

121      GAG GGC AGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC  
121      L L K L E L W P Q L R W L P A D L A A F A

181      GAG GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC  
181      V R A L C C K R A L L A R A L A A A A A

241      GAG GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC  
241      D P E G F E G G C S L A W R L A E L A Q

301      GAG GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC  
301      Q R A A E T F L I E G S R R F S Y S E A

361      GAG GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC  
361      E R E S N R A A R A F L R A L G W D W G

421      GAG GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC  
421      P D G G D S G E G S A G E G E R A A F G

481      GAG GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC  
481      A G D A A A G S G A E F A G G G G A A R

541      GAG GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC  
541      G G G E F A A F L S F G A T V A L L L F

601      GAG GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC  
601      A G P E F L W L W F G L A R A G L R T A

661      GAG GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC  
661      F V P T A L R R G P L L H C L R S C G A

721      GAG GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC  
721      R A L V L A P E F L E E L E P L L F A L

781      GAG GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC  
781      R A M G L R L W A A G F G T R F A G I S

841      GAG GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC  
841      D L L A E V E A E V D G P V F G Y L S S

901      GAG GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC  
901      P Q S E T D T C L Y I F T S G T T G L F

961      GAG GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC  
961      R A A R I S H L R I L C C Q G F Y G L C

1021      GAG GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC  
1021      G V E Q E D V E F L A L P L Y H M E G S

1081      GAG GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC  
1081      L L G L V G C M G E G A T V V L R S K F

1141      GAG GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC  
1141      S A G Q F W E D C Q Q E R V T V F Q Y E

1201      GAG GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC GGC  
1201      G E L C R Y L V N Q P F S R A E R G H X

Figure 94A

1261 ggc cgc cgc gca ggc ggc agc ggc cgc cgc cca gac acc tgg gag cgc ttc ggc cgc cgc  
425 V R L A V G S G L R P D T W E R F V R R

1321 acc ggc cgc cgc cgc ggc ggc gag acc tac gga cgc acc gag ggc acc ggc gcc acc acc  
435 F G P L Q V L E T Y G L T E G N V A T I

1381 acc tac acc gga cgc cgc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc  
455 N Y T G Q R G A V G R A S W L Y R H I F

1441 cgc ttc cgc cgc acc cgc tac gac ggc acc acc gga gag cca acc cgc gac acc cgc ggc  
475 P F S L I R Y D V T T G E F I R D P Q G

1501 cgc cgc acc ggc acc cgc cgc ggc ggc gag cca ggc cgc ggc ggc ggc ggc ggc ggc ggc  
495 H C M A T S P G E P G L L V A P V S Q Q

1561 cgc cgc cgc cgc ggc cgc ggc ggc ggc cgc gag cgc ggc ggc ggc ggc ggc ggc ggc ggc  
515 S P F L G Y A G G P E L A Q G K L L K D

1621 ggc ttc cgc cgc ggc ggc gac ggc ttc ttc acc acc ggc gac cgc cgc ggc ttc gac gac cca  
535 V F R F G D V F F N T G D L L V C D D Q

1681 ggc cgc cgc cgc ttc cgc gac cgc acc gga gac acc cgc agc cgc agc ggc ggc acc ggc  
555 G F L R F E D R T G D T F R W R G E N V

1741 ggc acc acc gag ggc ggc gag ggc ttc gag ggc cca gac ttc cgc cgc gag ggc acc ggc  
575 A T T E V A E V F E A L D F L Q E V N V

1801 cgc gga ggc acc ggc cgc ggc cgc gga ggc agc ggc gga acc ggc cca ggc ggc ggc ggc  
595 Y G V T V P G H E G R A G M A A L V L R

1861 cgc cgc cgc ggc ttc ggc cgc cgc cgc cgc acc cgc ggc ttc gag acc cgc cgc cgc  
615 F P E A L D L M C L Y T E V S E N L F P

1921 cgc ggc ggc cgc cgc ttc cgc agc cgc cgc gag ggc ttc ggc acc acc gag acc cgc acc  
635 Y A R P R F L R L Q E S L A T T E T F K

1981 cgc cgc acc ggc cgc agc ggc acc gag ggc cgc gac cgc agc acc cgc cgc gac cgc ggc  
655 Q Q K V R M A N E G F C P S T L S D F L

2041 cgc ggc cgc gac cgc ggc ggc ggc ggc tac cgc cgc cgc acc acc ggc cgc cgc agc ggc  
675 Y V L D Q A V G A Y L P L T T A R Y E A

2101 cgc cgc gga gga acc cgc cgc acc tgc gga cgc cgc cgc cgc agc cgc agc gag gga  
695 L L A G N L R C P

2161 cgc cgc

Figure 94B

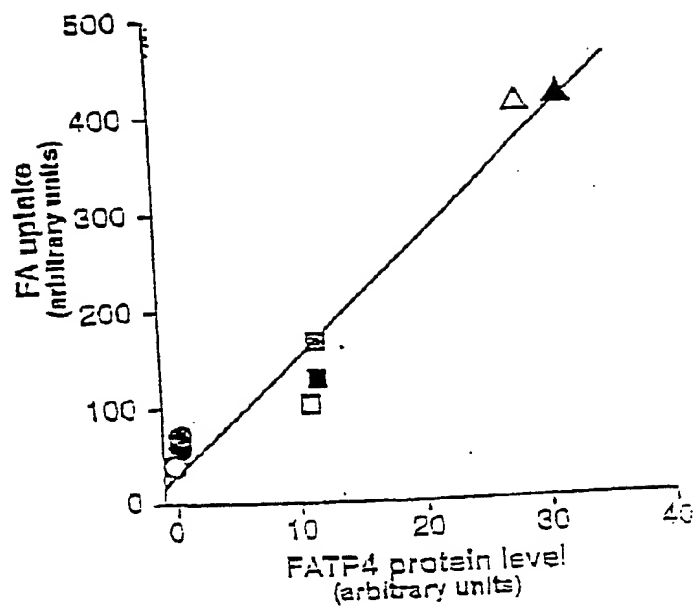


Figure 93

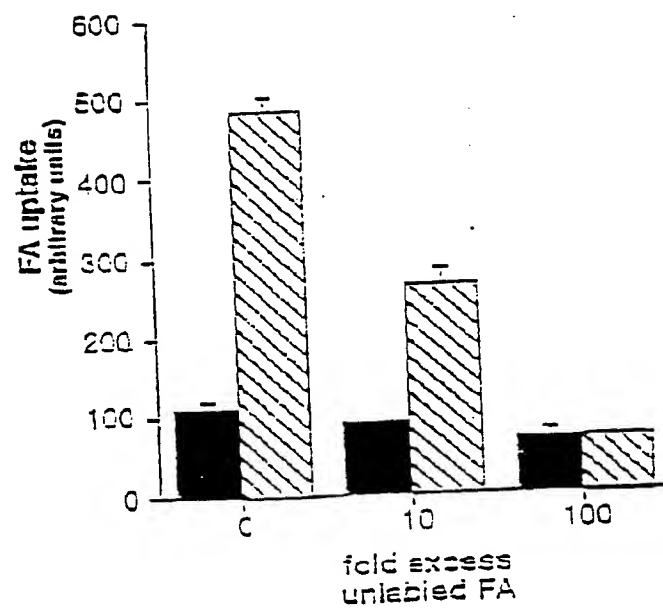


Figure 9b



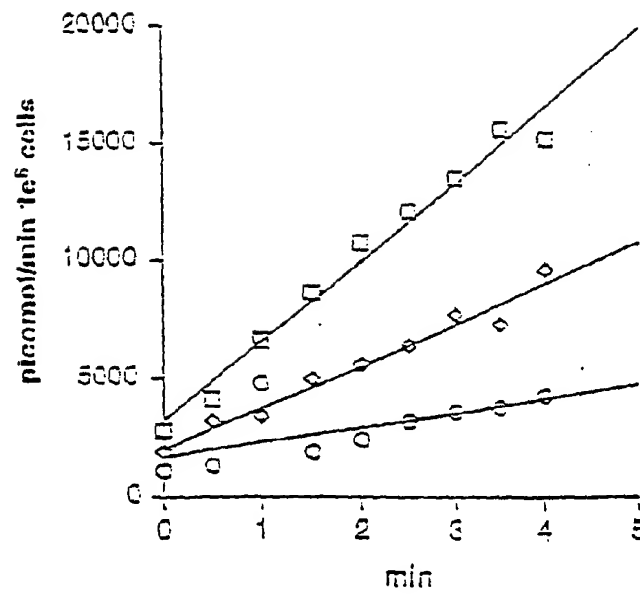


Figure 97

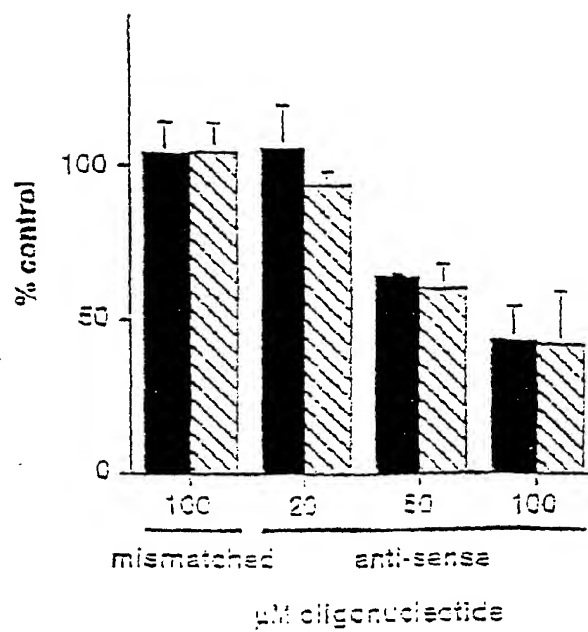


Figure 88

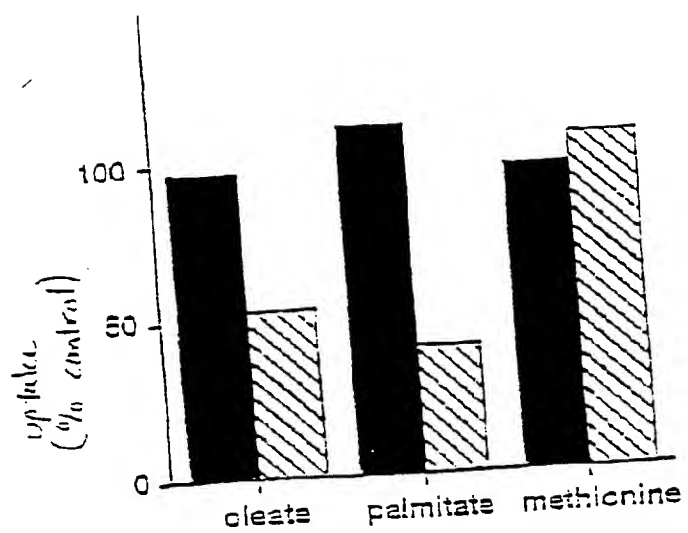


Figure 99

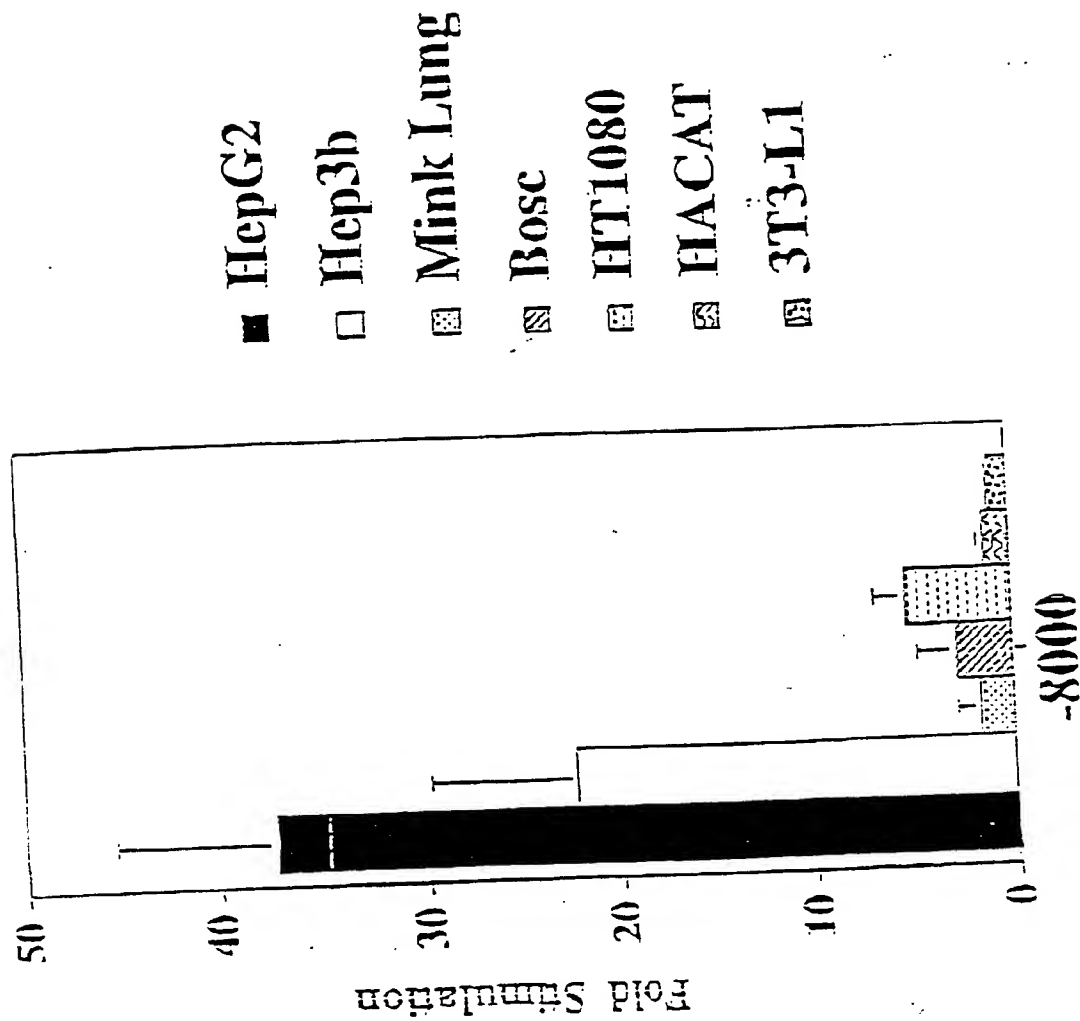


Figure 100

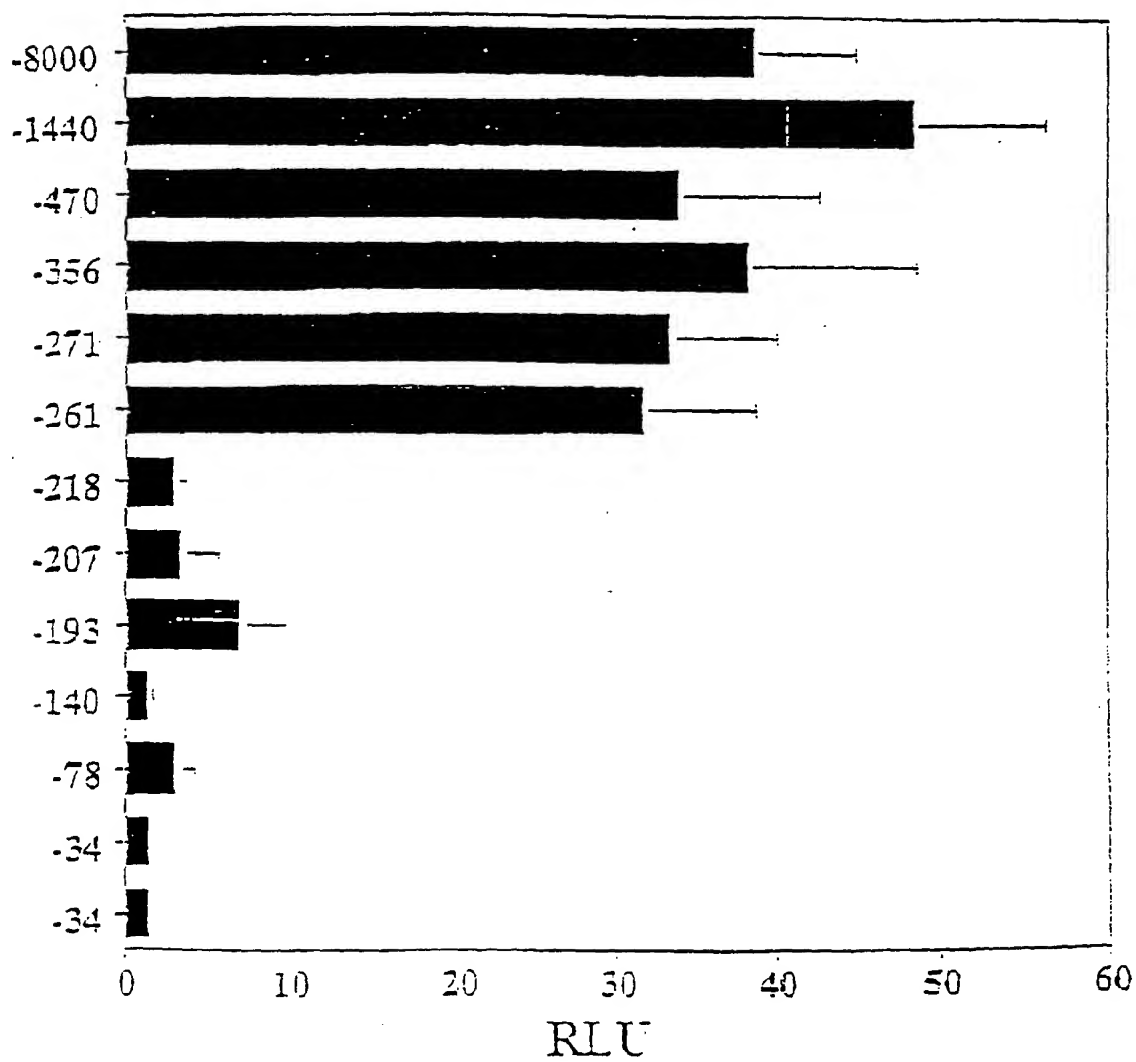
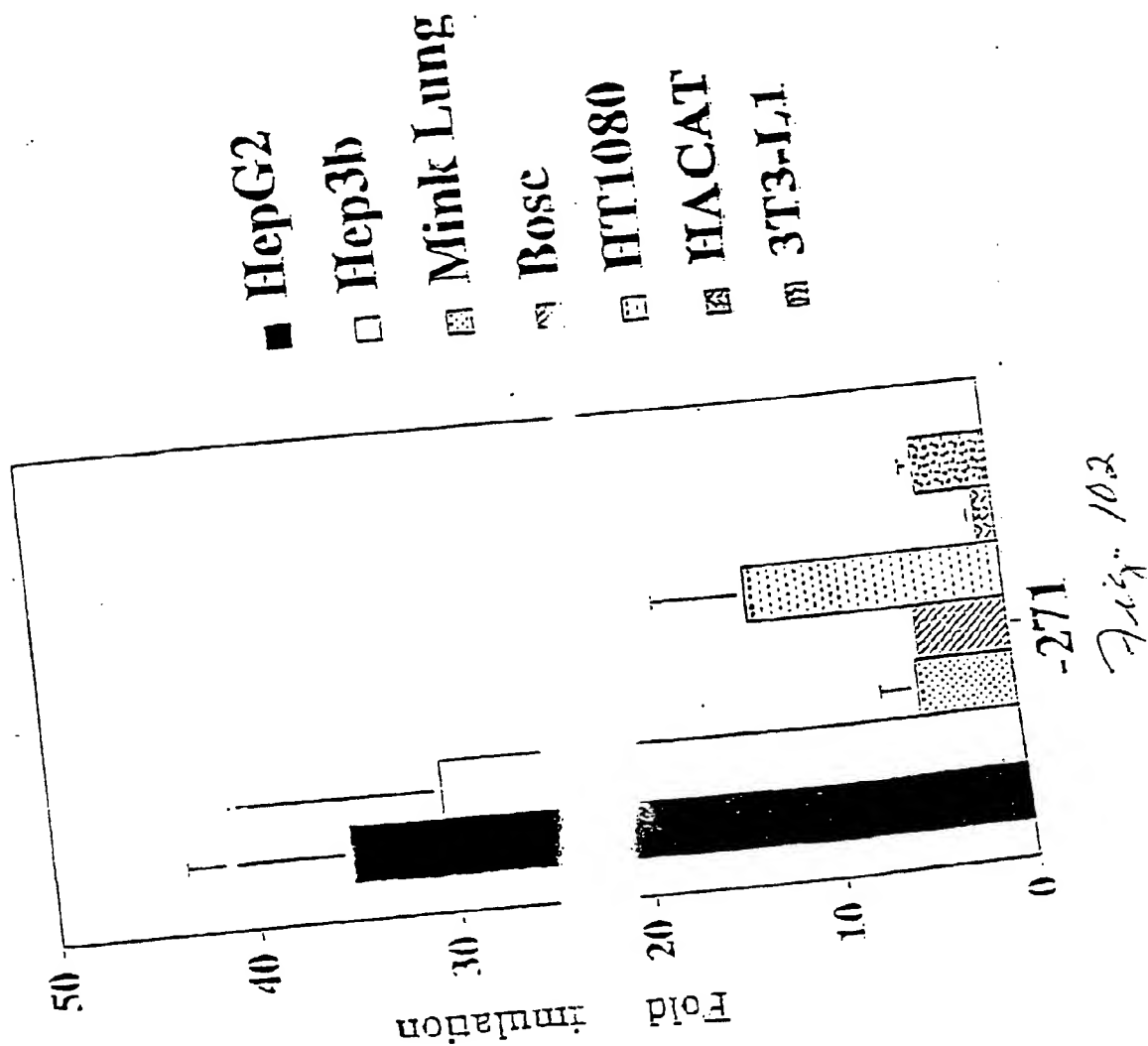


Fig. 101



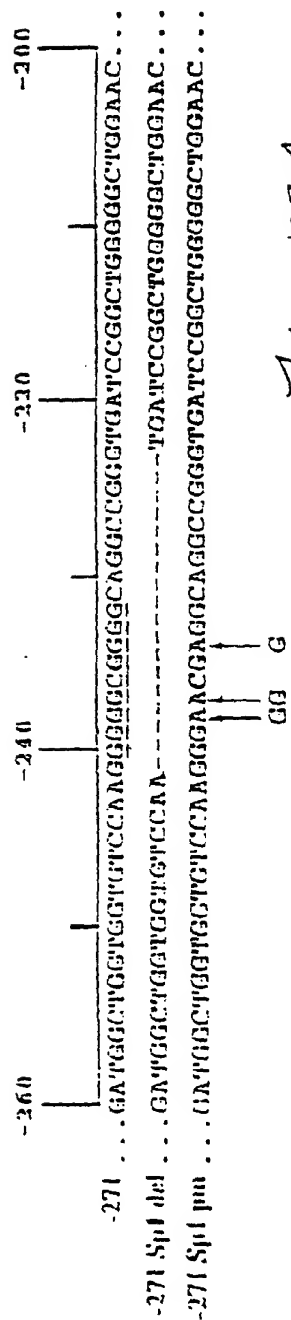


Fig. 103A

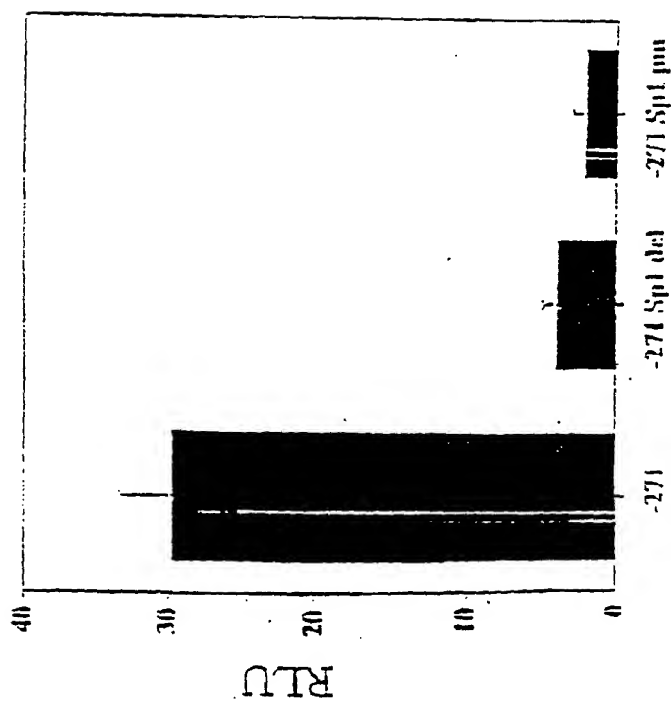
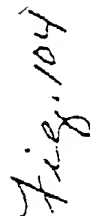
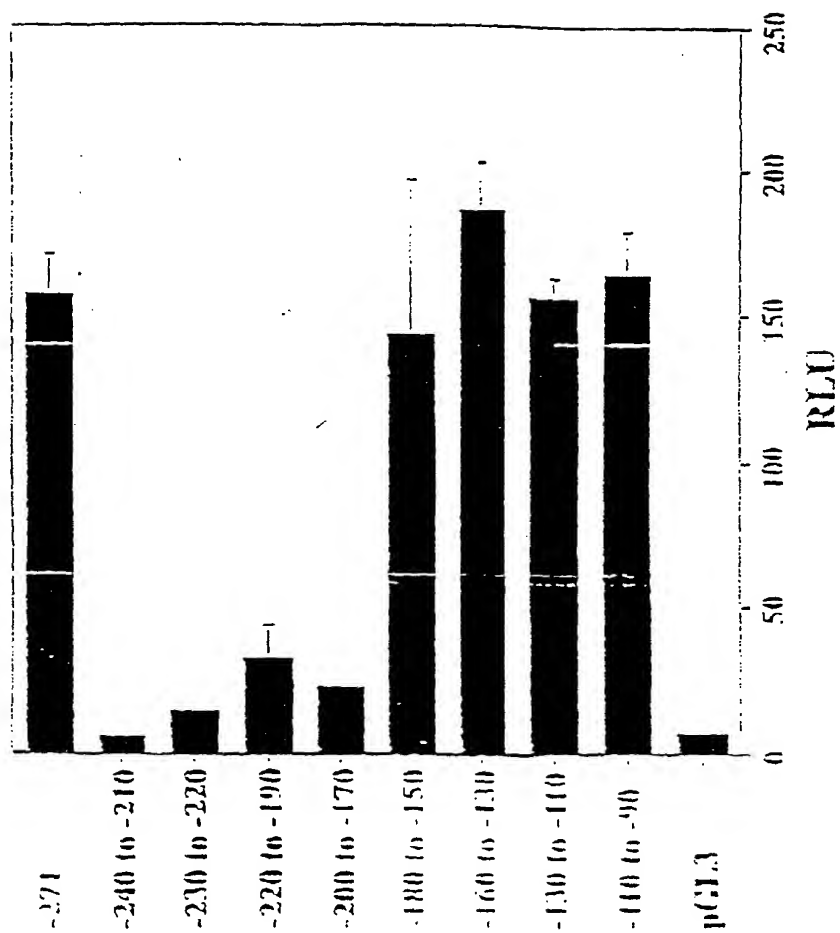
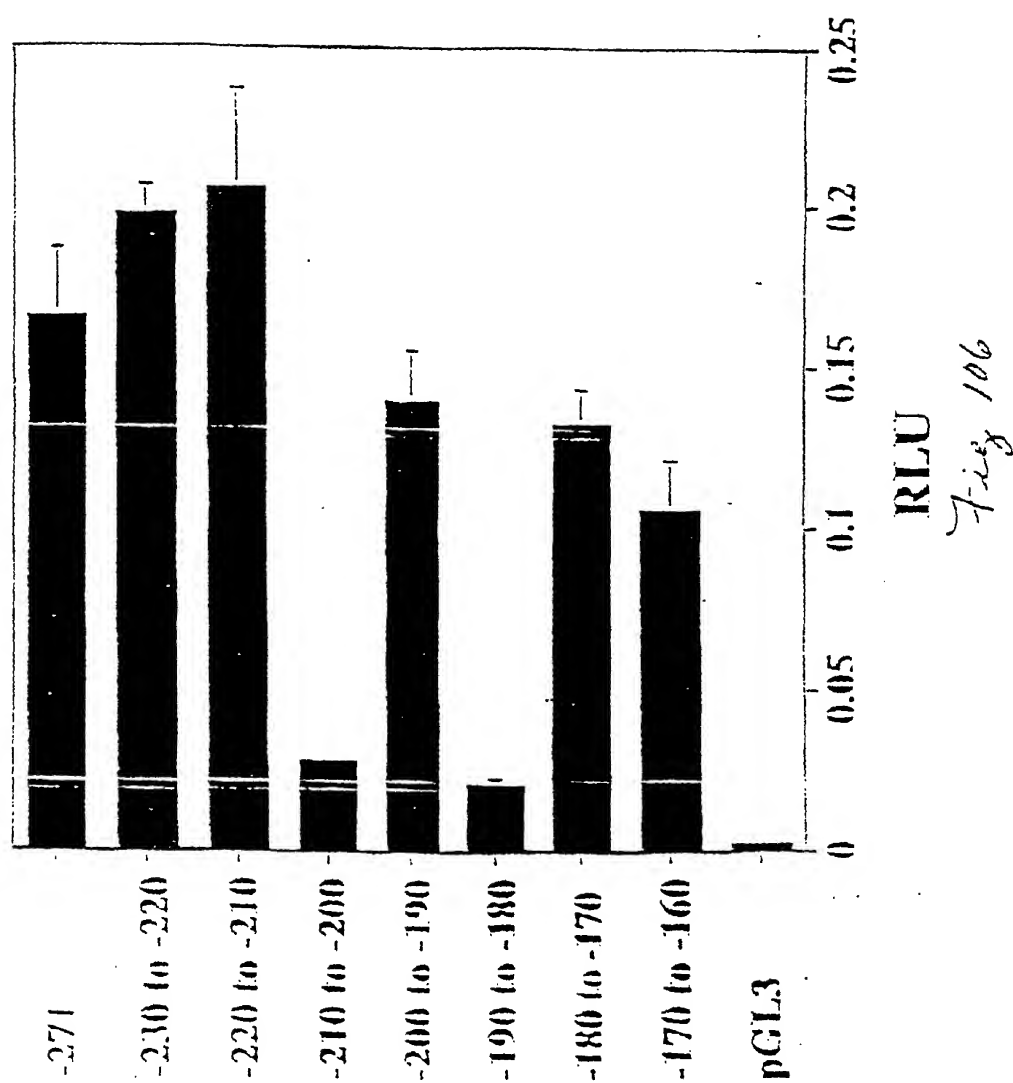


Fig. 103B



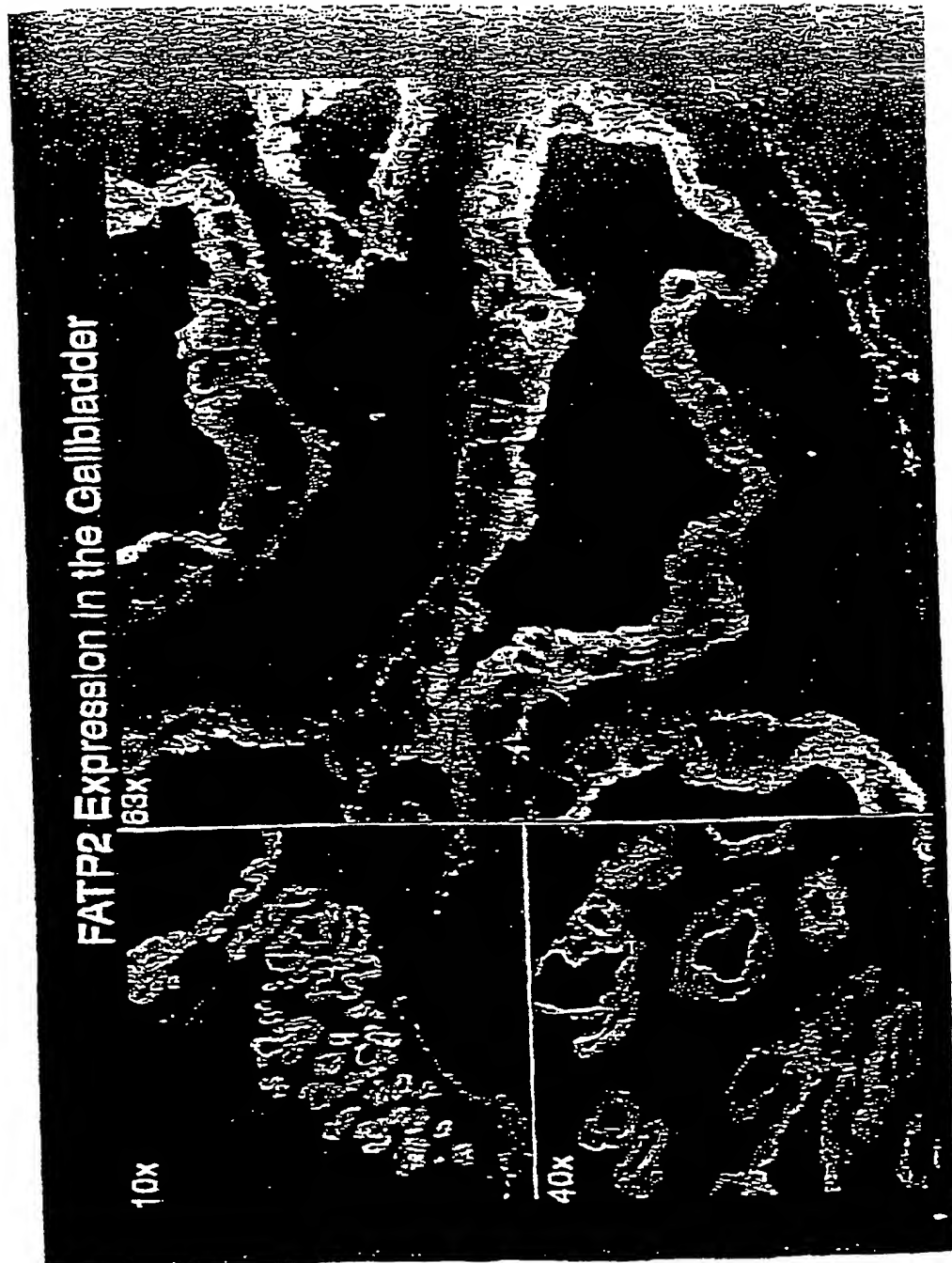


*Fig. 105*

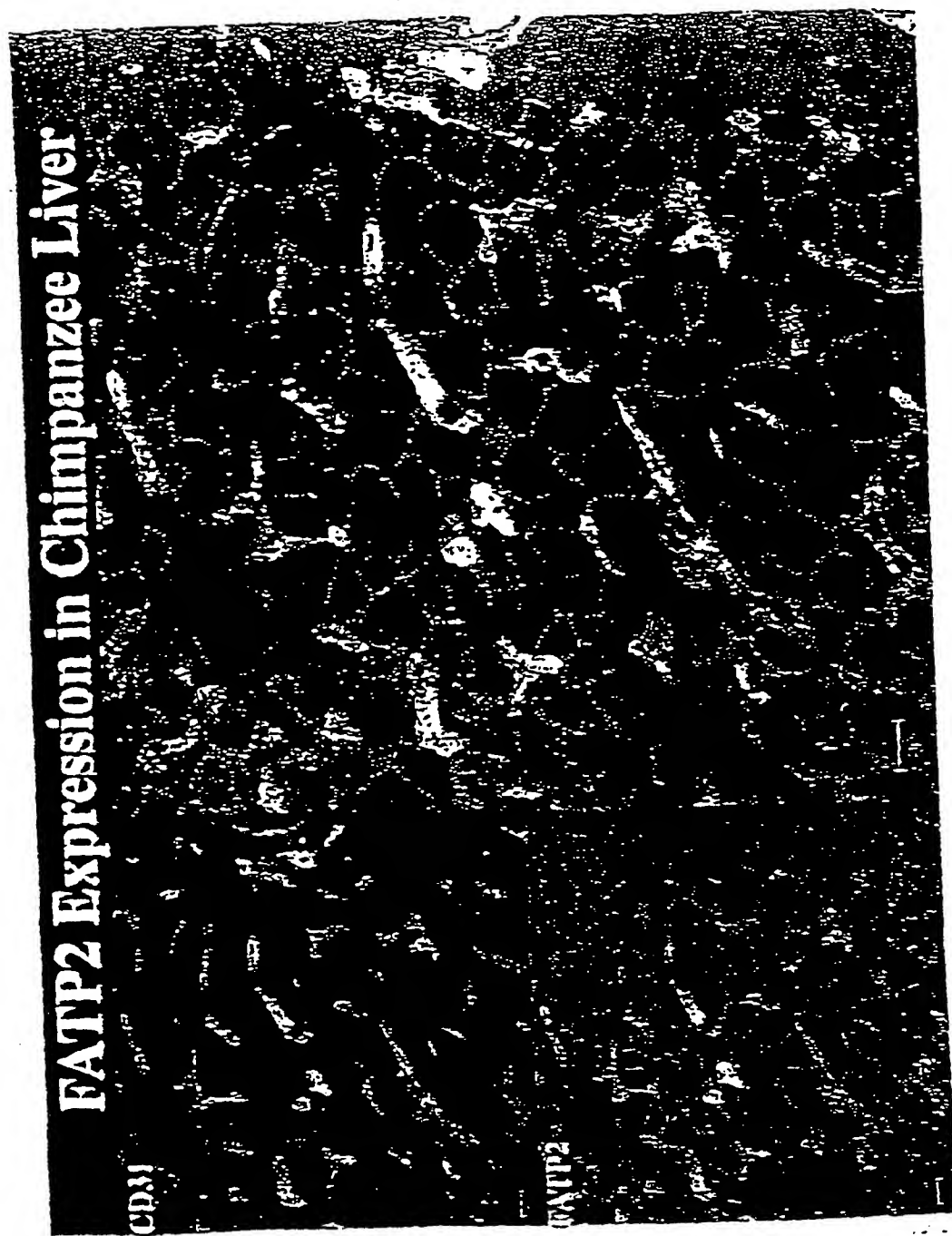


GAGGATGGCTGGTGGTGTCCAAGGGGGGCGGGGGA  
GGCCGGGTGATCCGGCTGGGGGGCTGGGAAGTGTAG  
AATTCCAGCCAGTAAGAACTAAGTAACAAAAGG  
ACAGAGTCCATGGGTCACATTCAGTTGCTGATAG  
TACTTGGTCATATTTGGGAAGTGGGTAGACAGAT  
TTCCTTAAAGGCAGGTAGTTAGGGCTTTGGAGCA  
CTCATCAGAGCTAAGAGAGATTACACGCTCTCAT  
CTACTTCAGAAAGAGCCAATGCCATG→

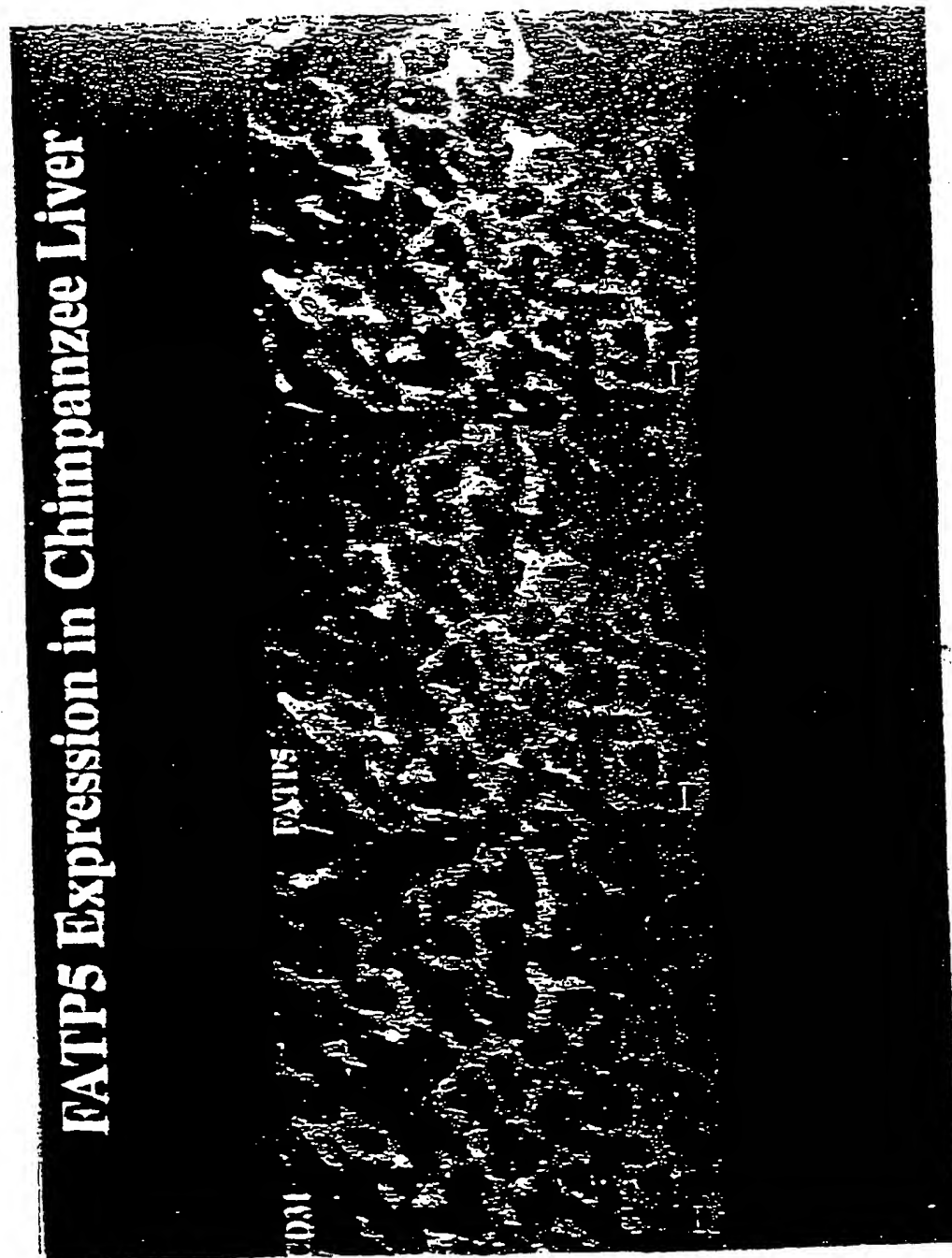
*Fig. 157*



*Figure 108*



*Figure 109*



*Figure 110*

## HsFATP3 (Johni003f04):

CCCGGGTTTCTGCTCTCCGCCCCGTGTGGAGTGGTGGGGCCCTGGGTGGGAATGGGCGTGTGCCAGCGCACGG  
GCGCTCCC  
TGGAAAGGAGAAGTCTCAGCTAGAACGAGCGGCCCTAGGTTTTTCGGAAGGGAGGATCAGGGATGTTTTGCCAGC  
GGCTGGAA  
CCAGACGGTGCCGATAGAGGAAGCGGGCTCCATGGGTGCCCTCCTGCTGTGCCCCCTGCTGCTGTTGCTACC  
GCTGCTGC  
TGCTGAAGCTACACCTCTGGCCGCGATTGCGCTGGCTTCGGGCGGACTTGCCCTTTGCGGTGCGAGCTCTGT  
GCTGCAAA  
AGGGCTCTTCGAGCTCGCGCCCTGGCCGCGGTGCGGCCGACCCGGAAGGTCCCGAGGGGGGCTGCAGCCTG  
GCCTGGCG  
CCTCGCGAACTGGCCGAGCAGCGCGCCGCGCACACCTTTCTCATTACGGCTCGCGGCGCTTTAGCTACTC  
AGAGGCGG  
AGCGCGAGAGTAAACAGGCTGCACGCGCCTTCTACCTGCGCTAGGCTGGGACTGGGGACCCGACGGCGGGC  
ACAGCGGC  
GAGGGAGCGCTGAGAGAAGGCGAGCGGGCAGCGCCGGGAGCCGGAGATGACGCGCCGGAAGCGCGCGGAG  
TTTGCCGG  
AGGGGACGGTGCCGCCAGAGGTGGAGGAGCCGCGCCCTCTGTACCTGGAGCACTGTGGCGCTGCTCCT  
CCCCGCTG  
GCCCAGAGTTTTCTGTGGCTCTGGTTCCGGCTGGCCAGGCGCGCCTGCGCACTGCGCTTTGTGCCACCGCCC  
TGCGCCGG  
GGCCCCCTGCTGCACTGCGCTCCGCGAGCTGCGGGCGCGCGCGCGCTGGTGTGCGCGCCAGAGTTTCTGGAGTCC  
CTGGAGCC  
GGACCTGCGCCCGCCCTGAGAGCCATGGGGCTCCACCTGTGGGCTGCAGGCCAGGAACCCACCTGCTGGAAT  
TAGCGATT  
TGCTGGCTGAAGTGTCCGCTGAAGTGGATGGGCGAGTGCCAGGATACCTCTCTCCCCCAGAGCATTAACAG  
ACACGTGC  
CTGTACATCTTACCTCTGGCACCAAGCGCCCTCCCGAAGGCTGCTCGGATCAGTCACTGAAAGATCCTGCAA  
TGCCAGGG  
CTTCTATCAGCTGTGTGGTGTCCACCAGGAAGATGTGATCTACCTCGCCCTCCCACTCTACCACATGTCCGG  
TTCCCTGC  
TGGGCACTCGGGCTGCATGGGCAATGGGGCCACAGTGGTGTGAAATCCAGTTCTCGGCTGGTCACTTCT  
GGGAAGAT  
TGCCAGCAGCACAGGGTGACGGTGTTCAGTACATTGGGGAGCTGTGCCGATACCTTGTCAACAGCCCGCG  
AGCAAGGC  
AGAACGTGGCCATAAGGTCCGGCTGGCAGTGGGCGAGCGGGCTGCGCCGAGATACCTGGGAGCGTTTTGTGG  
GCGCTTGG  
GGCCCCCTGAGGTGCTGGAGACATATGGACTGACAGAGGGCCAGCTGGCCACCATCACTACACAGGACAGC  
GGGGCGCT  
GTGGGGCTGCTTCCTGGCTTTACAAGCATATCTTCCCTTCTCCTTGATTGCTATGATGTCAACACAGGA  
GAGCCAA  
TCGGGACCCCCAGGGCACTGTATGGCCACATCTCCAGGTGAGCCAGGGCTGCTGGTGGCCCCCGGTAAAGCA  
GCAGTCCC  
CATTCCTGGGCTATGCTGGCGGGCCAGAGCTGGCCGAGGGGAAGTTGCTAAAGCACTCTTCCGGCCTGGGG  
ATGTTTTT  
TTCAACACTGGGGACCTGCTGGTCTGCGATGACCAAGCTTTTTCTCCGCTTCCATGATCGTACTGGAGACACC  
TTCAGGTG  
GAGGGGGAGAAATGTGGCCACACCGAGGTGGCAGAGGTCTTCGAGGCCCTAGATTTTCTTCAGGAGGTGAA  
CTCTATG  
GAGTCACTGTGCCAGGGCATGAAGGCAGGGCTGGAATGGCAGCCCTAGTTCTGCGTCCCCCCCCACGCTTTGG  
ACCTTATG

FIG. 111A

CAGCTCTACACCCACGTCCTCTGAGAACTTGCACCTTATGCCCGCCCGATTCTCTCAGGCTCCAGGAGTCT  
TTGGCCAC  
CACAGAGACCTTCAAACAGCAGAAAAGTTCCGATGCGAAATGAGGGCTTCGACCCACGACCCCTGTCTGACCC  
ACTGTACG  
TTCTGGACCAGGCTGTAGGTGCCTACCTGCCCTCACTGCGCGGTACAGCGCCCTCCTGGCAGGAAACC  
TTGGAATC  
TGAGAACTTCCACACCTGAGGCACTGAGAGAGGAACTCTGTGGGTGGGGCCGTTGCGGTGTACTGGGC  
TGTGAGG  
ATCTTTTCTATACCGAAGCTGCCGTCACTATTTGTAAATAAATGTGGCTGAGCTGATCCAGCTGTCTCTGA  
AAAAAAA  
AA

FIG. 111B



## FATP3 protein:

MGVCCQRTAPWKZYSQLZRAALGFRRKGGSGMFASGWNQTVPIZEAGSMRAALLLLPLLLPLLLPLLLKLHLWEQ  
LRWLPADL  
APAVRALCCXRALRARALAAAADPEGPEGGCSLAWRLAELAQQRAAETYLIEGSRRFSTYSZAZRESNRAAR  
AFLRALGW  
DWGFLGGDSGEGSAGEGZRAAPCAGDAAAGSGASTAGCGGAARGCGAAAPLSPGATVALLLPAGPEFLWLWF  
GLAKGLR  
TAPVFTALRRGPILLECLRSCCARALVLAPEFLZSLFPLPALRAMGLELWAAGPGTEFAGISDLLAEVSAEV  
DGFVPGYL  
SSPQSITDTCLYIFTSGTGLPKAARISELXILQCGFYQLCGVEQEDVITLALPLYEMSGSLLGITVGCMT  
GATVVLAS  
KFSAGQFWEDCQQRVTVFYIGELCRFLVNQPPSKAERGEXVRLAVGSGLRPDTHRFVFRFGPIQVLETY  
GLTEGNVA  
TINYTGQRGAVGRASWLYKSLPPFSLIRYDVTGEPIDPQGECMATSPGEFGLLVAFVSQQSPFLGYAGGP  
ELAQQKLL  
KDVFRPGDVFFNTCDLLVCDQGTFLRFEDRTGDTFRKXGENVATTEVAEVFEALDFLQEVNVYGVTVFGZEG  
RAGMAALV  
LRFFEALDLMQLYTEVSENLPFYARPRFLRLQESLATTITFKQKVRMANEGFDPSTLSDPLYVLDQAVGAY  
LPLTTARY  
SALLAGNLRI

FIG. 112

BODIPY-FA uptake in cos cells

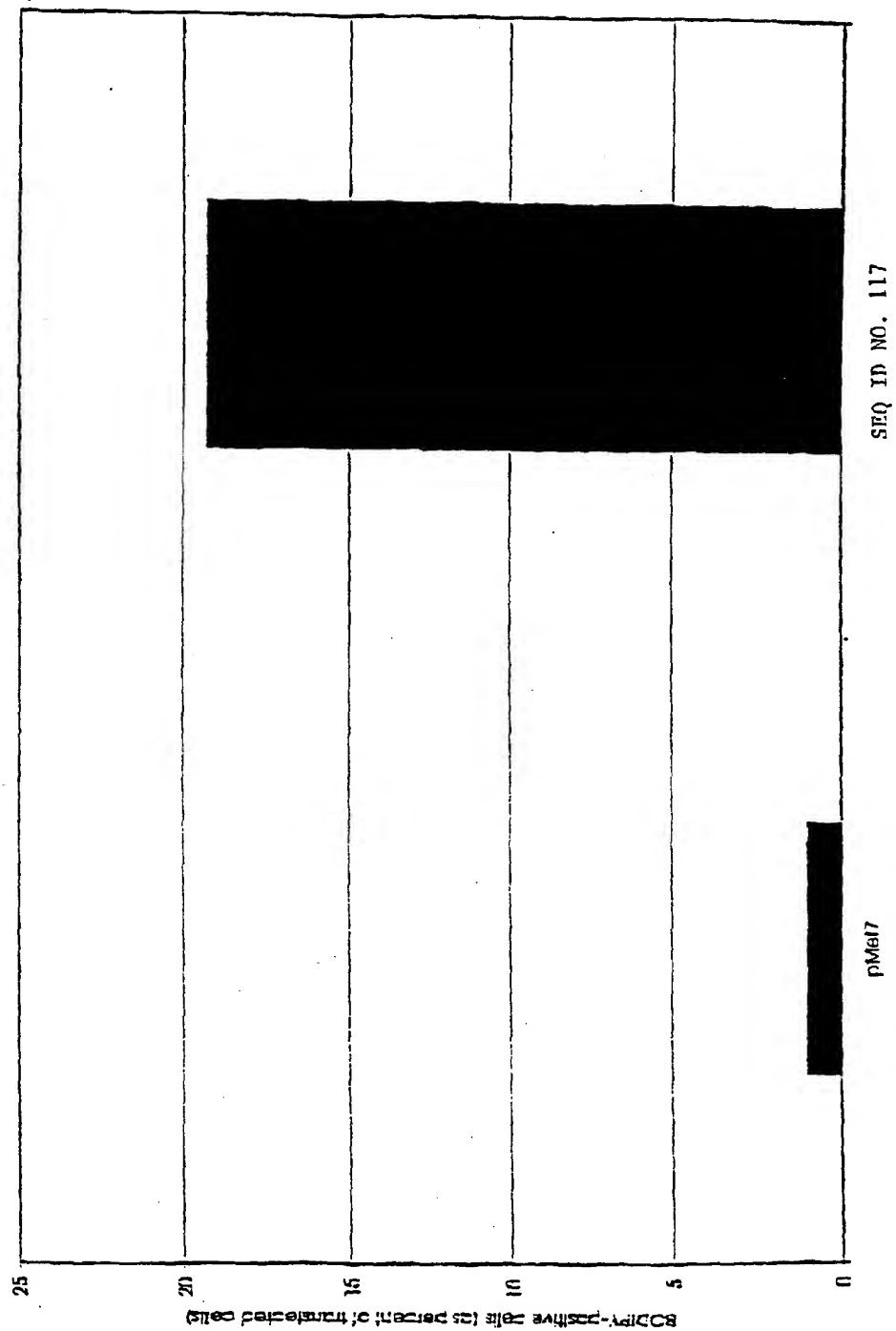


Fig. 113

## Peptide Sequence of hsFATP fragments for fatty acid binding experiments:

SP1:

RVFIKTIIRDIFGGLVLLKVKAKVRQCLQERRTVPIFASTVRRHPDKTALIFE  
 GTDTHWTFRQLDEYSSSVANFLQARG  
 LASGDVAAIFMENRNEFVGLWLGMAKLGVEAALINTNLRRDALLHCLTTSRAR  
 ALVFGSEMASAICEVHASLDPSLSLFC  
 SGSWEPGAVPPSTEHLDPILLKDAKHLPSCPDKGFTD

Fig. 114A

SP2:

RVFIKTIIRDIFGGLVLLKVKAKVRQCLQERRTVPIFASTVRRHPDKTALIFE  
 GTDTHWTFRQLDEYSSSVANFLQARG  
 LASGDVAAIFMENRNEFVGLWLGMAKLGVEAALINTNLRRDALLHCLTTSRAR  
 ALVFGSEMASAICEVHASLDPSLSLFC  
 SGSWEPGAVPPSTEHLDPILLKDAKHLPSCPDKGFTDKLFYTYTSGTTGLPKAAIV  
 VHSRYRMAALVYYGFRMRPNDIV  
 YDCLPLYH

Fig. 114B

SP3:

GDVAAIFMENRNEFVGLWLGMAKLGVEAALINTNLRRDALLHCLTTSRARALV  
 FGSEMASAICEVHASLDPSLSLFCSGS  
 WEPGAVPPSTEHLDPILLKDAKHLPSCPDKGFTDKLFYTYTSGTTGLPKAAIVVH  
 SRYRMAALVYYGFRMRPNDIVYDC  
 LPLYH

Fig. 114C

SP5:

RLVRVNEDTMELIRGPDGVCIPCOPGEPGQLVGRHQKDPLRRFDGYLNQGANNK  
 KLAQDVFKKGDQAYLTGDVIVMDEL  
 GYL YFRDRTGDTFRWKGENVSTTEVEGTL SRLLDMAADVAVYGVEVPGTEGRAC  
 MAAVASPTGNCDLERFAQVLEKELPLY  
 ARPFLRLPELEKTGTYKFKQTELKKEGFDPAIVKDPLFYLDQAQGRYVPLDQE  
 AYSRIQAGEEKL

Fig. 114D

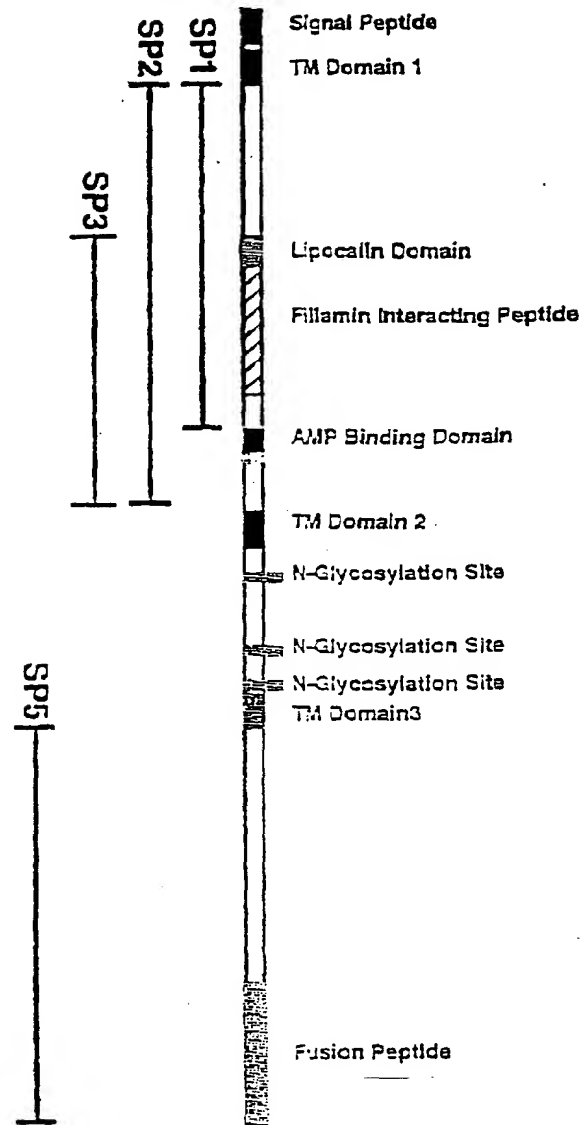


Fig. 115.

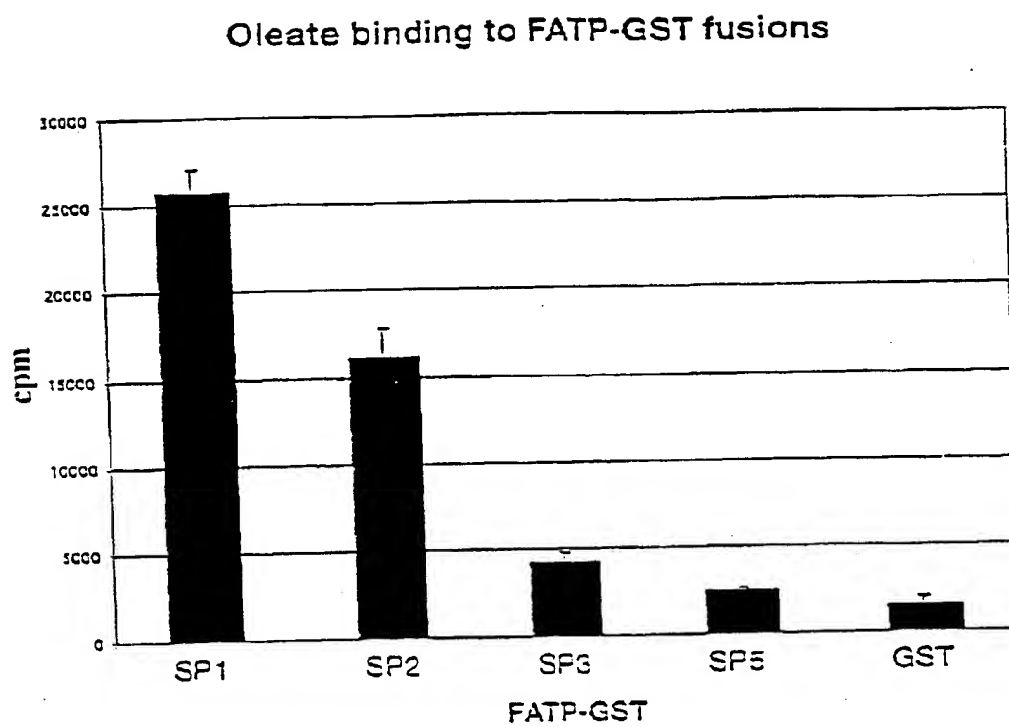


Fig. 116

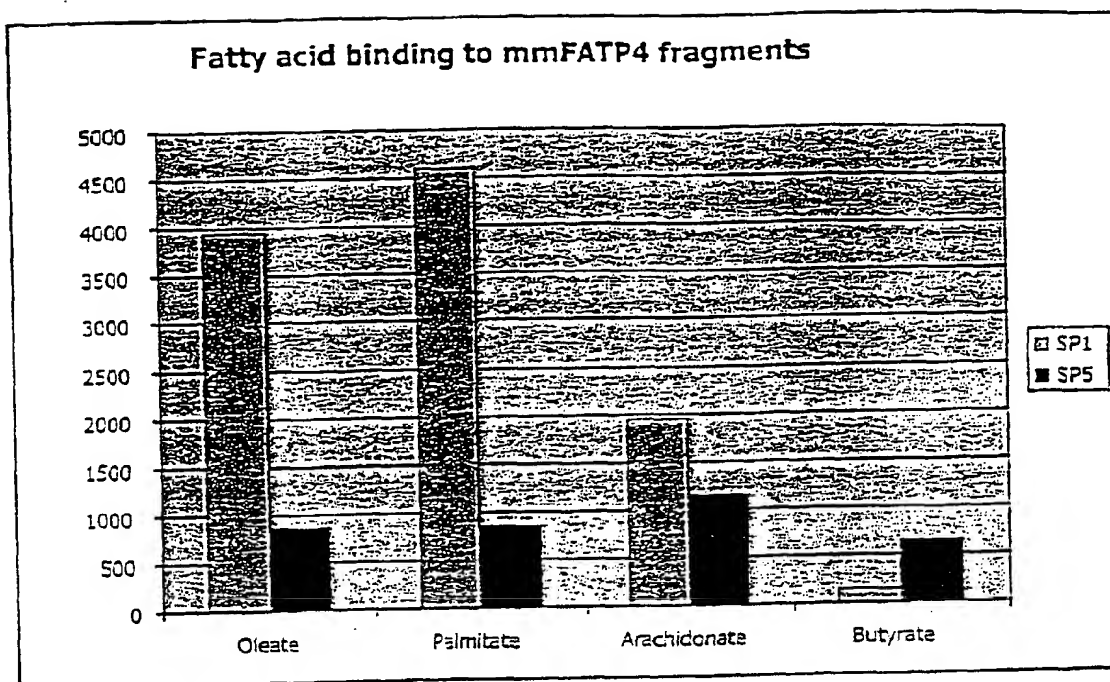


FIG. 117

Fig 118A

hsFATP6..	58	V	L	D	K	F	L	S	H	A	K	R	Q	P	R	K	P	F	L	I	V	E	G	D	-	-	-	I	Y	I	Y	Q	D	V	85	
hsFATP2..	55	I	L	R	A	F	L	E	K	A	R	Q	I	H	K	P	F	L	L	F	R	D	E	-	-	-	-	I	L	T	Y	A	Q	V	84	
hsFATP3..	114	L	A	W	R	L	A	E	L	A	Q	Q	H	A	A	H	T	F	L	I	H	G	S	R	-	-	-	-	R	E	S	Y	S	E	A	143
hsFATP5..	116	F	V	D	A	F	E	R	R	A	R	A	Q	P	G	H	A	L	L	V	W	T	G	P	G	A	G	S	V	T	F	G	E	L	148	
hsFATP4..	76	V	P	I	L	F	A	S	T	V	R	R	H	P	D	K	T	A	L	I	F	E	G	T	D	-	-	T	H	W	T	F	R	Q	L	107
hsFATP1..	70	I	P	R	I	F	Q	A	I	V	V	Q	R	Q	P	E	R	L	A	L	V	D	A	G	T	G	-	E	C	W	I	F	A	Q	L	109
hsFATP6..	86	D	K	R	S	S	R	V	A	H	V	E	L	N	H	S	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	101
hsFATP2..	85	D	R	R	S	N	Q	V	A	R	A	L	H	D	H	L	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100
hsFATP3..	144	E	R	E	S	N	F	A	A	R	A	F	L	H	A	L	G	W	D	W	G	P	D	G	G	D	S	G	E	G	S	A	G	E	-	176
hsFATP5..	149	D	A	R	A	G	Q	A	A	W	A	L	K	A	E	L	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	164
hsFATP4..	108	D	E	Y	S	S	S	V	A	N	-	F	L	Q	A	R	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	122
hsFATP1..	110	D	A	Y	S	N	A	V	A	N	-	L	F	R	Q	L	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	124
hsFATP6..	102	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	101	
hsFATP2..	101	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100	
hsFATP3..	177	G	E	R	A	P	G	A	G	D	A	A	G	S	G	A	E	F	A	G	G	D	G	A	A	R	G	G	G	A	A	A	A	A	-	209
hsFATP5..	165	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	167
hsFATP4..	123	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	122
hsFATP1..	125	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	124

Fig 118B



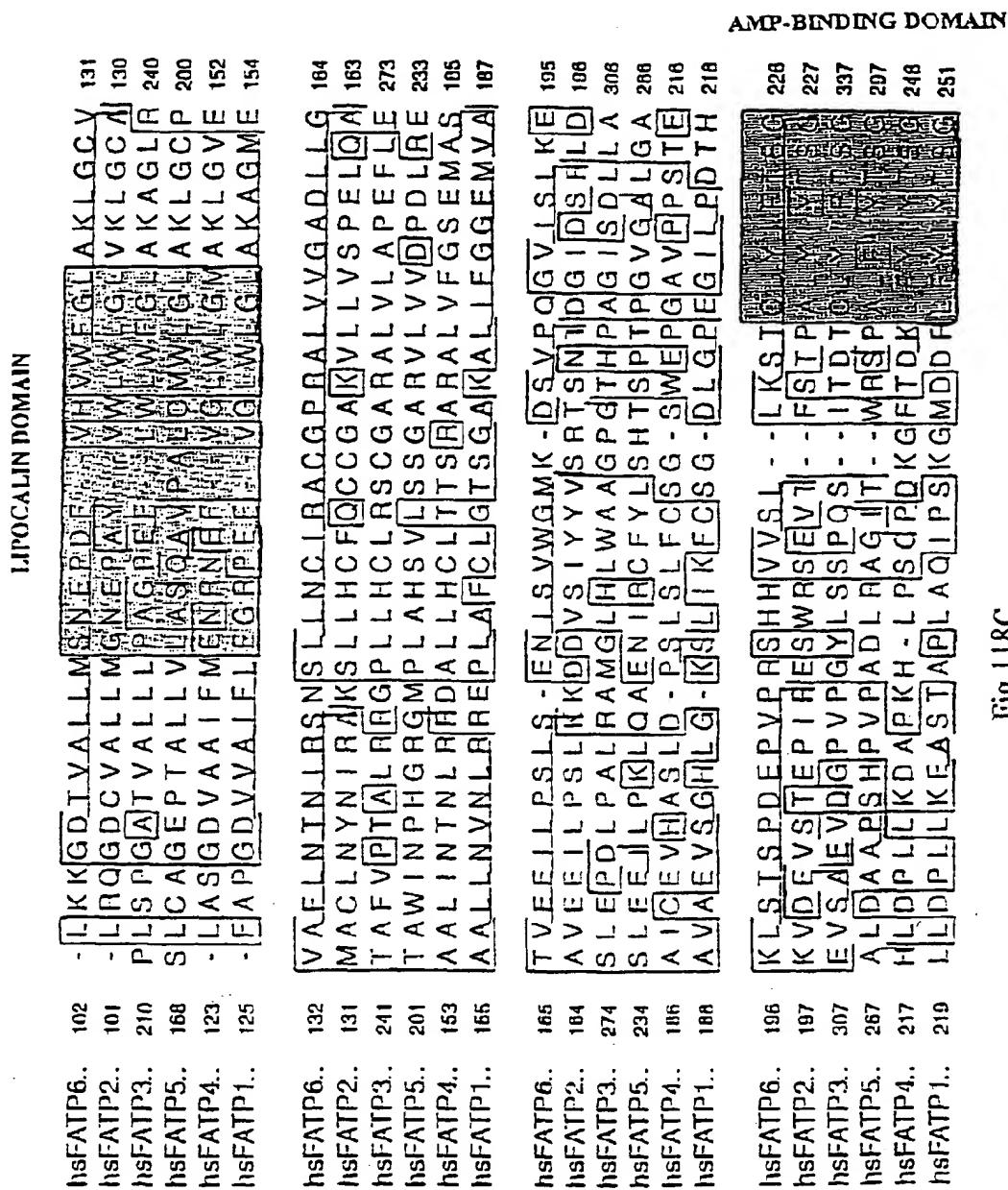


Fig 118C

AMP-BINDING DOMAIN  
CONTINUED

hsFATP6..	227	ITGLPKAAVLSQ	QL	QVLRGSAVLWAFG	CTAHD	258					
hsFATP2..	228	ITGLPKAAVLSQ	QL	IWYGTGLTFVSG	LKADDV	259					
hsFATP3..	338	ITGLPKAAVLSQ	QL	KILQCQGFYQLCG	VHQEDV	369					
hsFATP5..	238	ITGLPKAAVLSQ	QL	RVLQMSKMLSLSG	ATADDV	329					
hsFATP4..	249	ITGLPKAAVLSQ	QL	RYRMAALVYYGFR	RRND	281					
hsFATP1..	252	ITGLPKAAVLSQ	QL	RYRMAAEGHHAY	RRMQAADV	284					
hsFATP6..	259	VYITLLPLYHSSA	ALLG	ISGCVELGATC	VLLKKKE	291					
hsFATP2..	260	IYITLPFYHSA	ALLG	IHGCI	VAGATLALRTKF	292					
hsFATP3..	370	IYLAALPLYHMSGSL	LGI	VGCMGI	GATVVLKSKF	402					
hsFATP5..	330	VYTVLPPLYHVMGL	VVGI	LGCLDL	GATCVLAPKF	362					
hsFATP4..	282	VYDCLPLYHSA	GN	I	VGI	GQCL	LHGMTVVI	314			
hsFATP1..	285	LYDCLPLYHSA	GN	I	LGVGQCL	I	NGLTVVLRKKE	317			
hsFATP6..	292	SASQFEWS	DC	KKYD	VT	VFEQY	IGELCRYLCKQS	KR	324		
hsFATP2..	293	SASQFWD	DD	IRKYN	VT	VIQYIGEL	LRYLCNS	RQK	325		
hsFATP3..	403	SAGQFWE	DC	QQHR	VT	VFQYIGEL	CRYLVND	PPS	435		
hsFATP5..	363	STSCFW	DD	DC	RQHG	VT	VI	L	VVGEL	LRYLCNI	395
hsFATP4..	315	SASRFWD	DC	I	KYN	CT	I	VQYIGEL	CRYLLND	PPR	347
hsFATP1..	318	SASRFWD	DC	I	KYN	CT	I	VQYIGEL	CRYLLND	KQPV	350

Fig 118D

hsFATP6..	325	E	G	E	K	D	H	K	V	R	L	A	I	G	N	G	L	R	S	D	V	W	R	E	F	L	D	R	E	G	N	I	K	V	357
hsFATP2..	328	P	N	D	R	D	H	K	V	R	L	A	I	G	N	G	L	R	G	D	V	W	R	Q	F	V	K	R	F	G	D	I	C	I	358
hsFATP3..	438	K	A	E	R	G	H	K	V	R	L	A	V	G	S	G	L	R	P	D	T	W	E	R	F	V	R	R	F	G	P	L	Q	V	488
hsFATP5..	386	P	E	D	R	I	H	T	V	R	L	A	M	G	N	G	L	R	A	D	V	W	E	T	F	Q	Q	R	F	G	P	I	R	I	428
hsFATP4..	348	E	A	E	N	Q	H	Q	V	R	M	A	L	G	N	G	L	R	Q	S	I	W	T	N	F	S	S	R	F	H	I	P	Q	V	380
hsFATP1..	351	E	A	E	R	R	H	R	V	R	L	A	V	G	N	G	L	R	P	A	I	W	E	E	F	T	E	R	E	G	V	R	Q	I	383
hsFATP6..	358	C	E	L	Y	A	A	T	E	S	I	S	E	M	N	Y	I	G	R	I	G	A	I	G	R	T	N	L	E	Y	K	L			390
hsFATP2..	359	Y	E	F	Y	A	A	T	E	D	N	I	G	F	M	N	Y	A	R	K	V	G	A	V	G	R	V	N	Y	L	Q	K	K	I	301
hsFATP3..	469	L	E	T	Y	G	L	T	E	G	N	V	A	T	I	N	Y	T	G	Q	R	G	A	V	G	R	A	S	W	L	Y	K	H	I	501
hsFATP5..	429	W	E	V	Y	G	S	T	E	G	N	M	G	L	V	N	Y	V	G	R	D	G	A	L	G	K	M	S	C	L	L	R	M	L	481
hsFATP4..	381	A	E	F	Y	G	A	T	E	C	N	C	S	L	G	N	F	D	S	Q	V	G	A	C	G	F	N	S	R	I	L	S	F	V	413
hsFATP1..	384	G	E	F	Y	G	A	T	E	C	N	C	S	L	A	N	M	D	G	K	V	G	S	C	G	F	N	S	H	L	L	P	H	V	416
hsFATP6..	391	S	T	F	D	L	I	K	Y	D	F	Q	K	D	E	P	M	R	N	E	Q	G	W	C	I	H	V	K	K	G	E	P	G	L	423
hsFATP2..	392	I	T	Y	D	L	I	K	Y	D	V	E	K	D	E	P	V	R	D	E	N	G	Y	C	V	R	V	P	K	G	E	V	G	L	424
hsFATP3..	502	H	P	F	S	L	I	R	Y	D	V	T	T	G	E	P	I	R	D	P	Q	Q	H	C	M	A	T	S	P	G	E	P	G	L	534
hsFATP5..	482	S	P	F	E	L	V	Q	F	D	M	E	A	N	E	P	V	R	D	N	Q	G	F	C	I	P	V	G	L	G	E	P	G	L	484
hsFATP4..	414	Y	P	I	R	L	V	R	V	N	E	D	T	M	E	L	I	R	G	P	D	G	V	C	I	P	C	Q	P	G	E	P	G	Q	446
hsFATP1..	417	Y	P	I	R	L	V	K	V	N	E	D	T	M	E	L	L	R	D	A	D	G	L	C	I	P	C	Q	A	G	E	P	G	L	449
hsFATP6..	424	L	I	S	R	V	N	A	K	N	P	E	F	G	Y	A	G	-	-	P	Y	K	H	I	K	D	K	L	C	D	V	E	K	454	
hsFATP2..	425	L	V	C	K	I	T	Q	L	I	P	F	N	G	Y	A	G	-	-	A	K	A	Q	T	E	K	K	L	R	D	V	F	N	455	
hsFATP3..	535	L	V	A	P	V	S	Q	Q	S	P	F	L	G	Y	A	G	-	-	G	P	E	L	A	Q	G	K	L	K	D	V	F	R	565	
hsFATP5..	495	L	L	T	K	V	S	Q	Q	P	F	V	G	Y	R	G	-	-	P	R	E	L	S	E	H	K	L	V	R	N	V	R	Q	525	
hsFATP4..	447	L	V	G	R	I	I	Q	K	D	P	L	R	R	E	D	G	Y	L	N	Q	G	A	N	N	K	K	I	A	K	D	V	F	K	479
hsFATP1..	450	L	V	G	Q	I	N	Q	Q	D	P	L	R	R	F	D	G	Y	V	S	E	S	A	T	S	K	K	L	A	H	S	V	F	S	482

Fig 118E

hsFATP6..	455	K	G	D	V	Y	L	N	I	G	D	L	L	V	Q	D	N	E	L	Y	F	W	D	R	I	G	D	I	F	R	W	K	487		
hsFATP2..	456	K	G	D	L	Y	F	N	S	G	D	L	L	M	V	D	H	E	N	F	I	Y	H	H	D	R	V	G	D	T	F	R	W	K	488
hsFATP3..	566	P	G	D	V	F	F	N	T	G	D	L	L	V	C	D	Q	G	F	L	R	F	H	D	R	T	G	D	T	F	R	W	K	598	
hsFATP5..	526	S	I	G	D	V	Y	N	T	G	D	V	L	A	M	D	R	E	I	F	L	Y	F	R	D	R	L	G	D	T	F	R	W	K	558
hsFATP4..	480	K	G	D	Q	A	Y	L	T	G	D	V	L	V	M	D	E	L	G	Y	L	Y	F	R	D	R	T	G	D	T	F	R	W	K	512
hsFATP1..	403	K	G	D	S	A	Y	L	S	G	D	V	L	V	M	D	E	L	G	Y	M	Y	E	R	D	R	S	G	D	T	F	R	W	R	516
hsFATP6..	488	G	E	N	V	A	T	T	E	V	A	D	V	I	G	M	L	D	F	I	Q	E	A	N	V	Y	G	V	A	L	S	G	Y	F	520
hsFATP2..	489	G	E	N	V	A	T	T	E	V	A	D	T	V	G	L	V	D	F	V	Q	E	V	N	V	Y	G	V	H	V	P	D	H	E	521
hsFATP3..	509	G	E	N	V	A	T	T	E	V	A	E	V	F	E	A	L	D	F	L	Q	E	V	N	V	Y	G	V	T	V	P	G	H	E	531
hsFATP5..	559	G	E	N	V	S	T	H	E	V	E	G	V	L	S	Q	V	D	F	L	Q	Q	V	N	V	Y	G	V	C	V	P	G	C	E	591
hsFATP4..	513	G	E	N	V	S	T	T	E	V	E	G	T	L	S	R	L	L	D	M	A	D	V	A	V	Y	G	V	E	V	P	G	T	E	545
hsFATP1..	516	G	E	N	V	S	T	T	E	V	E	G	V	L	S	R	L	L	D	Q	T	D	V	A	V	Y	G	V	A	V	P	G	V	E	548
hsFATP6..	521	G	R	A	G	M	A	S	I	L	L	K	P	N	T	S	L	D	L	E	K	V	E	Q	V	V	T	E	L	P	A	Y	A	553	
hsFATP2..	522	G	R	I	G	M	A	S	I	K	M	K	E	N	H	E	F	D	G	K	K	L	F	Q	H	I	A	D	Y	L	P	S	Y	A	554
hsFATP3..	532	G	R	A	G	M	A	A	L	V	L	R	P	P	H	A	L	D	L	M	Q	L	Y	T	H	V	S	E	N	L	P	P	Y	A	564
hsFATP5..	592	G	K	V	G	M	A	A	V	Q	L	A	P	G	Q	T	F	D	G	E	H	L	Y	Q	H	V	R	A	W	L	P	A	Y	A	624
hsFATP4..	546	G	R	A	G	M	A	A	V	A	S	-	P	T	G	N	C	D	L	E	R	F	A	Q	V	L	E	K	E	L	P	L	Y	A	577
hsFATP1..	540	G	K	A	G	M	A	A	V	A	D	-	P	H	S	L	L	D	P	N	A	I	Y	Q	E	L	Q	K	V	L	A	P	Y	A	590

Fig 118F

hsFATP6..	554	C	P	R	F	L	R	I	Q	E	K	M	E	A	T	G	I	E	K	L	L	K	H	Q	L	V	E	D	G	F	N	P	L	K	586		
hsFATP2..	555	R	P	R	F	L	R	I	Q	D	T	I	E	I	T	G	T	F	K	H	R	K	M	T	L	V	E	E	G	F	N	P	A	V	587		
hsFATP3..	585	F	P	R	F	L	R	I	Q	E	S	L	A	T	T	E	T	F	K	Q	K	V	R	M	A	N	E	G	F	D	P	S	T	587			
hsFATP5..	525	T	P	H	F	I	R	I	Q	D	A	M	E	V	T	S	T	F	K	L	M	K	T	R	L	V	R	E	G	F	N	V	G	I	587		
hsFATP4..	578	R	P	I	F	L	R	L	P	E	L	H	K	T	G	T	Y	K	F	Q	K	T	E	L	R	K	E	G	F	D	P	A	I	581			
hsFATP1..	581	F	P	I	F	L	R	L	P	Q	V	D	T	I	G	I	F	K	I	Q	K	T	R	L	Q	R	E	G	F	D	P	R	Q	581			
hsFATP6..	587	I	S	E	P	L	Y	F	E	M	D	N	L	K	K	S	Y	Y	L	L	T	R	E	L	Y	D	Q	I	M	L	G	E	L	K	L	587	
hsFATP2..	588	I	K	D	A	L	Y	F	L	D	D	T	A	K	M	Y	V	M	P	M	T	I	E	D	I	Y	N	A	I	S	A	K	T	L	K	L	588
hsFATP3..	598	L	S	D	P	L	Y	V	L	D	Q	A	V	G	A	Y	L	P	L	T	T	A	R	Y	S	A	L	L	A	G	N	L	R	I	598		
hsFATP5..	650	V	V	D	P	L	F	V	L	D	N	R	A	Q	S	F	R	P	L	T	A	E	M	Y	Q	A	V	C	E	G	I	T	W	R	L	598	
hsFATP4..	611	V	K	D	P	L	F	Y	L	D	A	Q	K	G	R	Y	V	P	L	D	Q	E	A	Y	S	R	I	Q	A	G	E	E	K	L	598		
hsFATP1..	614	T	S	D	R	L	F	L	D	L	K	Q	G	H	Y	L	P	L	N	E	A	V	Y	T	R	I	C	S	G	A	F	A	L	598			

Fig 118C



(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
29 March 2001 (29.03.2001)

PCT

(10) International Publication Number  
**WO 01/21795 A3**(51) International Patent Classification<sup>7</sup>: **C12N 15/12**,  
C07K 14/705, C12N 5/10, 15/62, G01N 33/50, 33/68,  
C12Q 1/68, C07K 16/28, A61K 47/48 // A61P 3/04

(21) International Application Number: PCT/US00/25891

(22) International Filing Date:  
21 September 2000 (21.09.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
09/405,505 23 September 1999 (23.09.1999) US  
09/405,504 23 September 1999 (23.09.1999) US  
09/465,280 16 December 1999 (16.12.1999) US  
09/506,252 17 February 2000 (17.02.2000) US  
09/611,197 6 July 2000 (06.07.2000) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	09/405,505 (CIP)
Filed on	23 September 1999 (23.09.1999)
US	09/405,504 (CIP)
Filed on	23 September 1999 (23.09.1999)
US	09/465,280 (CIP)
Filed on	16 December 1999 (16.12.1999)
US	09/506,252 (CIP)
Filed on	17 February 2000 (17.02.2000)
US	09/611,197 (CIP)
Filed on	6 July 2000 (06.07.2000)

(71) Applicants (for all designated States except US): **WHITE-HEAD INSTITUTE FOR BIOMEDICAL RESEARCH** [US/US]; Nine Cambridge Center, Cambridge, MA 02142 (US). **MILLENNIUM PHARMACEUTICALS, INC.** [US/US]; 75 Sidney Street, Cambridge, MA 02139 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **STAHL, Andreas** [DE/US]; 12 Farrington Avenue, #2, Allston, MA 02134 (US). **HIRSCH, David, J.** [US/US]; 17 Rockview Street, #3, Jamaica Plain, MA 02130 (US). **LODISH, Harvey, F.** [US/US]; 195 Fisher Avenue, Brookline, MA 02446 (US). **GIMENO, Ruth, E.** [DE/US]; 65 Beverly Road, Wellesley, MA 02481 (US). **TARTAGLIA, Louis, A.** [US/US]; 32 Manor House Road, Newton, MA 02459 (US).(74) Agents: **HOGLE, Doreen, M.** et al.; Hamilton, Brook, Smith & Reynolds, P.C., Two Militia Drive, Lexington, MA 02421 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report

(88) Date of publication of the international search report:  
22 November 2001

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: **FATTY ACID TRANSPORT PROTEINS**

(57) Abstract: A family of fatty acid transport proteins (FATPs) mediate transport of long chain fatty acids (LCFAs) across cell membranes into cells. These proteins exhibit different expression patterns among the organs of mammals. Nucleic acids encoding FATPs of this family, vectors comprising these nucleic acids, as well as the production of FATP proteins in host cells are described. Also described are methods to test FATPs for fatty acid transport function, and methods to identify inhibitors or enhancers of transport function. The altering of LCFA uptake by administering to the mammal an inhibitor or enhancer of FATP transport function of a FATP in the small intestine can decrease or increase calories available as fats, and can decrease or increase circulating fatty acids. The organ specificity of FATP distribution can be exploited in methods to direct drugs, diagnostic indicators and so forth to an organ such as the heart.

WO 01/21795 A3

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 00/25891

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> IPC 7 C12N15/12 C07K14/705 C12N5/10 C12N15/62 G01N33/50 G01N33/68 C12Q1/68 C07K16/28 A61K47/48 //A61P3/04		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b> Minimum documentation searched (classification system followed by classification symbols) IPC 7 C12N C07K		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used) EPO-Internal, BIOSIS		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 99 36537 A (TARTAGLIA LOUIS A ;WHITEHEAD BIOMEDICAL INST (US); GIMENO RUTH E () 22 July 1999 (1999-07-22) cited in the application	1-43,45, 47,49, 50, 52-66, 70-76, 79, 81-86, 88,90, 92,94, 96,119, 120, 123-125
Y	the whole document	70-86, 88,90, 92,94, 96, 122-125
--- -- -/--		
<div style="display: flex; justify-content: space-between;"> <span><input checked="" type="checkbox"/> Further documents are listed in the continuation of box C.</span> <span><input checked="" type="checkbox"/> Patent family members are listed in annex.</span> </div>		
<div style="display: flex;"> <div style="flex: 1;"> <p>* Special categories of cited documents :</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> </div> <div style="flex: 1;"> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>"Z" document member of the same patent family</p> </div> </div>		
Date of the actual completion of the international search  <div style="text-align: center; font-size: 1.2em;">15 February 2001</div>		Date of mailing of the international search report  <div style="text-align: center; font-size: 1.2em;">10.05.01</div>
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2, NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016		Authorized officer  <div style="text-align: center; font-size: 1.2em;">ANDRES S.M.</div>



## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/25891

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>FLOWER D: "The lipocalin protein family: structure and function"  BIOCHEMICAL JOURNAL,  vol. 318, 1996, pages 1-14, XP002095126  ISSN: 0264-6021  the whole document</p> <p>---</p>	<p>70-86,  88,90,  92,94,  96,  122-125</p>
X	<p>WO 99 46281 A (BAKER KEVIN P ;CHEN JIAN (US); GENENTECH INC (US); GURNEY AUSTIN ())  16 September 1999 (1999-09-16)</p> <p>abstract  page 10, paragraph 15.  page 275 -page 276; example 114  figure 39  claims</p> <p>---</p>	<p>3,31,  33-42,  55,61,  62,  70-76,  81-85</p>
A	<p>HIRSCH ET AL: "A family of fatty acid transporters conserved from mycobacterium to man"  PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA,  vol. 95, July 1998 (1998-07), pages  8625-8629, XP002107853  ISSN: 0027-8424  cited in the application</p> <p>---</p>	
A	<p>BANASZAK, L. ET AL.: "Lipid-binding proteins: A family of fatty acid and retinoid transport proteins."  ADVANCES IN PROTEIN CHEMISTRY,  vol. 45, 1994, pages 89-151, XP000982080  ISBN: 0-12-034245-6  cited in the application</p> <p>---</p>	
X	<p>DATABASE EM_EST [Online]  EMBL;  ID: HSAA13688; Accession number :  AA193416, 23 January 1997 (1997-01-23)  HILLIER, L. ET AL.: "zr40b06.r1  Soares NhHMPu S1 Homo sapiens cDNA clone  IMAGE:665843 5'"  XP002160453  abstract  &amp; HILLIER, L. ET AL.: "Generation and expression of 280,000 human expressed sequence tags"  GENOME RESEARCH,  vol. 6, 1996, pages 807-828, XP000914615</p> <p>---</p> <p>-/--</p>	<p>4-6,29,  30</p>

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/25891

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE EM_EST [Online]  EMBL;  Accession number : AI041230,  1 July 1998 (1998-07-01)  "ov56b02.x1 Soares testis NHT Homo  sapiens cDNA clone IMAGE:1641291 3'"  XP002160454  abstract</p> <p>-----</p>	<p>4-6,29,  30</p>

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US 00/25891

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
  
2. ☒ Claims Nos.: 44 46 48 51 87 89 91 93 95 97 121  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:  
see FURTHER INFORMATION sheet PCT/ISA/210
  
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
  
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
  
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
  
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:  
1-66, 119-121 (in totality) and 67-97, 122-125 (all partially)

### Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-66,119-121 (in totality) and 67-97,  
122-125 (all partially)

Human fatty acid transporter protein FATP3 characterised by SEQ IDs 116 and 117; fragments and variants thereof; fusion proteins, vectors or cells comprising them; antibodies against them; their use in methods for screening modulating agents, for diagnostic or therapy.

2. Claims: 67-97,122-125 (all partially)

As for subject 1, but concerning FATP1 characterised by SEQ IDs 46 and 47.

3. Claims: 67-97,101,122-125 (all partially) and 102-106,  
112-116 (totally)

As for subject 1, but concerning FATP2 characterised by SEQ IDs 48 and 49. Methods for directing an agent to liver cells or gall bladder using a moiety which binds to FATP2.

4. Claims: 67-97,122-125 (all partially)

As for subject 1, but concerning FATP4 characterised by SEQ IDs 52 and 53.

5. Claims: 67-97,101,122-125 (all partially) and 98-100,  
107-111,117-118 (totally)

As for subject 1, but concerning FATP5 characterised by SEQ IDs 54 and 55. A promoter sequence derived therefrom and a method for directing an agent to liver cells using a moiety which binds to FATP5.

6. Claims: 67-97,122-125 (all partially)

As for subject 1, but concerning FATP6 characterised by SEQ IDs 56 and 57.

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 44 46 48 51 87 89 91 93 95 97 121

Claims 44, 46, 48, 51, 87, 89, 91, 93, 95, 97 and 121 relate to compounds defined solely by a desired property or a method of identification. As there is no structural feature characterising these compounds and allowing a meaningful search, these claims have not been searched.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 00/25891

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 9936537	A	22-07-1999	AU 2310899 A	02-08-1999
			EP 1045904 A	25-10-2000
WO 9946281	A	16-09-1999	AU 3072199 A	27-09-1999
			AU 3075099 A	11-10-1999
			EP 1064382 A	03-01-2001
			WO 9947677 A	23-09-1999
			AU 1532499 A	15-06-1999
			EP 1032667 A	06-09-2000
			WO 9927098 A	03-06-1999
			AU 3757099 A	08-11-1999
			EP 1071773 A	31-01-2001
			WO 9954467 A	28-10-1999
			AU 1070399 A	10-05-1999
			EP 1025227 A	09-08-2000
			WO 9920756 A	29-04-1999

CORRECTED VERSION

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
29 March 2001 (29.03.2001)

PCT

(10) International Publication Number  
WO 01/21795 A3(51) International Patent Classification<sup>7</sup>: C12N 15/12,  
C07K 14/705, C12N 5/10, 15/62, G01N 33/50, 33/68,  
C12Q 1/68, C07K 16/28, A61K 47/48 // A61P 3/04(US). **HIRSCH, David, J.** [US/US]; 17 Rockview Street,  
#3, Jamaica Plain, MA 02130 (US). **LODISH, Harvey, F.**  
[US/US]; 195 Fisher Avenue, Brookline, MA 02446 (US).  
**GIMENO, Ruth, E.** [DE/US]; 65 Beverly Road, Welles-  
ley, MA 02481 (US). **TARTAGLIA, Louis, A.** [US/US];  
32 Manor House Road, Newton, MA 02459 (US).

(21) International Application Number: PCT/US00/25891

(22) International Filing Date:

21 September 2000 (21.09.2000)

(74) Agents: **HOGLE, Doreen, M.** et al.; Hamilton, Brook,  
Smith & Reynolds, P.C., 530 Virginia Road, P.O. Box 9133,  
Concord, MA 01742-9133 (US).

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

09/405,505	23 September 1999 (23.09.1999)	US
09/405,504	23 September 1999 (23.09.1999)	US
09/465,280	16 December 1999 (16.12.1999)	US
09/506,252	17 February 2000 (17.02.2000)	US
09/611,197	6 July 2000 (06.07.2000)	US

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,  
DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,  
HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,  
LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,  
NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,  
TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.(63) Related by continuation (CON) or continuation-in-part  
(CIP) to earlier applications:

US	09/405,505 (CIP)
Filed on	23 September 1999 (23.09.1999)
US	09/405,504 (CIP)
Filed on	23 September 1999 (23.09.1999)
US	09/465,280 (CIP)
Filed on	16 December 1999 (16.12.1999)
US	09/506,252 (CIP)
Filed on	17 February 2000 (17.02.2000)
US	09/611,197 (CIP)
Filed on	6 July 2000 (06.07.2000)

(84) Designated States (*regional*): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,  
IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG,  
CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report

(88) Date of publication of the international search report:  
22 November 2001(71) Applicants (*for all designated States except US*): **WHITE-  
HEAD INSTITUTE FOR BIOMEDICAL RESEARCH**  
[US/US]; Nine Cambridge Center, Cambridge, MA 02142  
(US). **MILLENNIUM PHARMACEUTICALS, INC.**  
[US/US]; 75 Sidney Street, Cambridge, MA 02139 (US).(48) Date of publication of this corrected version:  
30 May 2002(15) Information about Correction:  
see PCT Gazette No. 22/2002 of 30 May 2002, Section II

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **STAHL, Andreas**  
[DE/US]; 12 Farrington Avenue, #2, Allston, MA 02134For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

(54) Title: FATTY ACID TRANSPORT PROTEINS

(57) Abstract: A family of fatty acid transport proteins (FATPs) mediate transport of long chain fatty acids (LCFAs) across cell membranes into cells. These proteins exhibit different expression patterns among the organs of mammals. Nucleic acids encoding FATPs of this family, vectors comprising these nucleic acids, as well as the production of FATP proteins in host cells are described. Also described are methods to test FATPs for fatty acid transport function, and methods to identify inhibitors or enhancers of transport function. The altering of LCFA uptake by administering to the mammal an inhibitor or enhancer of FATP transport function of a FATP in the small intestine can decrease or increase calories available as fats, and can decrease or increase circulating fatty acids. The organ specificity of FATP distribution can be exploited in methods to direct drugs, diagnostic indicators and so forth to an organ such as the heart.

WO 01/21795 A3

## FATTY ACID TRANSPORT PROTEINS

## RELATED APPLICATION(S)

- This application is a continuation-in-part of U.S. Patent Application Number 09/506,252 filed February 17, 2000 which is a continuation-in-part of U.S.
- 5 Patent Application Number 09/465,280 filed December 16, 1999 which is a continuation-in-part of U.S. Patent Application Number 09/405,505 filed September 23, 1999, and is a continuation-in-part of U.S. Patent Application Number 09/232,195 filed January 14, 1999, both of which claim the benefit of U.S. Provisional Application Number 60/110,941 filed December 4, 1998; U.S.
- 10 Provisional Application Number 60/093,491 filed July 20, 1998; and U.S. Provisional Application Number 60/071,374 filed January 15, 1998. This application is also a continuation-in-part of U.S. Patent Application Number 09/405,504 filed September 23, 1999, which is a continuation-in-part of U.S. Patent Application Number 09/232,201 filed January 14, 1999, which claims the benefit of
- 15 U.S. Provisional Application Number 60/110,941 filed December 4, 1998; U.S. Provisional Application Number 60/093,491 filed July 20, 1998; and U.S. Provisional Application Number 60/071,374 filed January 15, 1998. This application is also a continuation-in-part of U.S. Patent Application Number 09/232,197 filed January 14, 1999, United States Patent Application Number
- 20 09/232,200 filed January 14, 1999 and International Application Number PCT/US99/00182 filed January 14, 1999, each of which claims the benefit of U.S. Provisional Application Number 60/110,941 filed December 4, 1998; U.S. Provisional Application Number 60/093,491 filed July 20, 1998; and U.S.



Provisional Application Number 60/071,374 filed January 15, 1998. The teachings of each of these referenced applications are incorporated herein by reference in their entirety.

#### GOVERNMENT SUPPORT

- 5                   The invention was supported, in whole or in part, by a grant from the National Heart, Lung, and Blood Institute (HL41484), by National Institutes of Health Grant DK 47618 and National Institutes of Health Grant 5 T32 CA 09541. The United States Government has certain rights in the invention.

#### BACKGROUND OF THE INVENTION

- 10               Long chain fatty acids (LCFAs) are an important source of energy for most organisms. They also function as blood hormones, regulating key metabolic functions such as hepatic glucose production. Although LCFAs can diffuse through the hydrophobic core of the plasma membrane into cells, this nonspecific transport cannot account for the high affinity and specific transport of LCFAs exhibited by
- 15               cells such as cardiac muscle, hepatocytes, enterocytes, and adipocytes. The molecular mechanisms of LCFA transport remains largely unknown. Identifying these mechanisms can lead to pharmaceuticals that modulate fatty acid uptake by the intestine and by other organs, thereby alleviating certain medical conditions (e.g. obesity).

#### 20   SUMMARY OF THE INVENTION

- Described herein is a diverse family of fatty acid transport proteins (FATPs) which are evolutionarily conserved; these FATPs are plasma membrane proteins which mediate transport of LCFAs across the membranes and into cells. Members of the FATP family described herein are present in a wide variety of organisms, from
- 25               mycobacteria to humans, and exhibit very different expression patterns in tissues among the organisms. FATP family members are expressed in prokaryotic and eukaryotic organisms and comprise characteristic amino acid domains or sequences which are highly conserved across family members. In addition, the function of the

FATP gene family is conserved throughout evolution, as shown by the fact that the *Caenorhabditis (C). elegans* and mycobacterial FATPs described herein facilitate LCFA uptake when they are overexpressed in COS cells or *Escherichia (E.) coli*, respectively. FATPs are expressed in a wide variety of tissues, including all tissues  
5 which are important to fatty acid metabolism (uptake and processing).

In specific embodiments, FATPs of the present invention are from such diverse organisms as humans (*Homo (H.) sapiens*), mice, (*Mus (M.) musculus*), *F. rubripes*, *C. elegans*, *Drosophila (D.) melanogaster*, *Saccharomyces (S.) cerevisiae*, *Aspergillus nidulans*, *Cochliobolu heterostrophus*, *Magnaporthe grisea* and  
10 *Mycobacterium (M.)*, such as *M. tuberculosis*. As described herein, four novel mouse FATPs, referred to as mmFATP2, mmFATP3, mmFATP4 and mmFATP5, and six human FATPs, referred to as hsFATP1, hsFATP2, hsFATP3, hsFATP4, hsFATP5 and hsFATP6, have been identified. All four novel murine FATPs (mmFATP2-5) and a previously identified murine FATP (renamed herein FATP1)  
15 have orthologs in humans (hsFATP1-5); the sixth human FATP (hsFATP6) does not as yet have a mouse ortholog. The expression patterns of these FATPs vary, as described in detail below.

The present invention relates to FATP family members from prokaryotes and eukaryotes, nucleic acids (DNA, RNA) encoding FATPs, and nucleic acids which  
20 are useful as probes or primers (e.g., for use in hybridization methods, amplification methods) for example, in methods of detecting FATP-encoding genes, producing FATPs, and purifying or isolating FATP-encoding DNA or RNA. Also the subject of this invention are antibodies (polyclonal or monoclonal) which bind an FATP or FATPs; methods of identifying additional FATP family members (for example,  
25 orthologs of those FATPs described herein by amino acid sequence) and variant alleles of known FATP genes; methods of identifying compounds which bind to an FATP, or modulate or alter (enhance or inhibit) FATP function; compounds which modulate or alter FATP function; methods of modulating or altering (enhancing or inhibiting) FATP function and, thus, LCFA uptake into tissues of a mammal (e.g.  
30 human) by administering a compound or molecule (a drug or agent) which increases or reduces FATP activity; and methods of targeting compounds to tissues by

administering a complex of the compound to be targeted to tissues and a component which is bound by an FATP present on cells of the tissues to which the compound is to be targeted. For example, a complex of a drug to be delivered to the liver and a component which is bound by an FATP present on liver cells (e.g., FATP5) can be  
5 administered.

In one embodiment, the present invention relates to modulating or altering (enhancing or inhibiting/reducing) LCFA uptake in the small intestine and, thus, increasing or reducing the number of calories in the form of fats available to an individual. In another embodiment, the present invention relates to inhibiting or  
10 reducing LCFA uptake in the small intestine in order to reduce circulating fatty acid levels; that is, LCFA uptake in the small intestine is reduced and, therefore, circulating (blood) levels are not as high as they otherwise would be. FATP4 has been shown to be expressed in epithelial cells of the small intestine and particularly in the brush border layer of the small intestine. FATP2 has also been shown to be  
15 expressed at low levels in epithelial cells of the small intestine, particularly in the duodenum. In contrast, FATP1, FATP3, FATP5 and FATP6 were not detected in any of the intestinal tissues. Thus, also described herein are FATPs which are present in the epithelial cell layer of the small intestine where they mediate LCFA uptake. These FATPs, particularly FATP4 and also FATP2, are targets for methods  
20 and drugs which block their function or activity and are useful in treating obesity, diabetes and heart disease. The ability of these FATPs to mediate fat uptake can be modulated or altered (enhanced or inhibited), thus modulating fat uptake in the small intestine. This can be done, for example, by administering to an individual, such as a human or other animal, a drug which blocks interaction of LCFAs with FATP4  
25 and/or FATP2 in the small intestine, thus inhibiting LCFA passage into the cells of the small intestine. As a result, fat absorption is reduced and, although the individual has consumed a certain quantity of fat, the LCFAs are not absorbed to the same extent they would have been in the absence of the compound administered.

Thus, one embodiment of this invention is a method of reducing LCFA  
30 uptake (absorption) in the small intestine and, as a result, reducing caloric uptake in the form of fat. A further embodiment is a compound (drug) useful in inhibiting or

reducing fat absorption in the small intestine. In another embodiment, the invention is a method of reducing circulating fatty acid levels by administering to an individual a compound which blocks interactions of LCFAs with FATP4 and/or FATP2 in the small intestine, thus inhibiting LCFA passage into cells of the small intestine. As a  
5 result, fatty acids pass into the circulatory system at a diminished level and/or rate, and circulating fatty acid levels are lower than they would be in the absence of the compound administered. This method is particularly useful for therapy in individuals who are at risk for or have hyperlipidemia. That is, it can be used to prevent the occurrence of elevated levels of lipids in the blood or to treat an  
10 individual in whom blood lipid levels are elevated. Also the subject of this invention is a method of identifying compounds which alter FATP function (and thus, in the case of FATP2 and/or FATP4, alter LCFA uptake in the small intestine).

In another embodiment, the present invention relates to a method of modulating or altering (enhancing or inhibiting) the function of FATP6, which is  
15 expressed at high levels in the heart. A method of inhibiting FATP6 function is useful, for example, in individuals with heart disease, such as ischemia, since reducing LCFA uptake into heart muscle in an individual who has ischemic heart disease, which may be manifested by, for example, angina or heart attack, can reduce symptoms or reduce the extent of damage caused by the ischemia. In this  
20 embodiment, a drug which inhibits FATP6 function is administered to an individual who has had or is having a heart attack, to reduce LCFA uptake by the individual's heart and, as a result, reduce the damage caused by ischemia. In a further embodiment, this invention is a method of targeting a compound, such as a therapeutic drug or an imaging reagent, to heart tissue by administering to an  
25 individual (e.g., a human) a complex of the compound and a component (e.g., a LCFA or LCFA-like compound) which is bound by an FATP (e.g., FATP6) present in cells of heart tissue.

In a further embodiment, LCFA uptake by the liver is modulated or altered (enhanced or reduced), in an individual. For example, a drug which inhibits the  
30 function of an FATP present in liver (e.g., FATP5) is administered to an individual

who is diabetic, in order to reduce LCFA uptake by liver cells and, thus reduce insulin resistance.

The present invention, thus, provides methods which are useful to alter, particularly reduce, LCFA uptake in individuals and, as a result, to alter (particularly  
5 reduce), availability of the LCFAs for further metabolism. In a specific embodiment, the present invention provides methods useful to reduce LCFA uptake and, thus, fatty acid metabolism in individuals, with the result that caloric availability from fats is reduced, and circulating fatty acid levels are lower than they otherwise would be. These methods are useful, for example, as a means of weight  
10 control in individuals, (e.g., humans) and as a means of preventing elevated serum lipid levels or reducing serum lipid levels in humans. FATPs expressed in the small intestine, such as FATP4, are useful targets to be blocked in treating obesity (e.g., chronic obesity) or to be enhanced in treating conditions in which enhanced LCFA uptake is desired (e.g., malabsorption syndrome or other wasting conditions).

15 The identification of this evolutionarily conserved fatty acid transporter family will allow a better understanding of the mechanisms whereby LCFAs traverse the lipid bilayer as well as yield insight into the control of energy homeostasis and its dysregulation in diseases such as diabetes and obesity.

#### BRIEF DESCRIPTION OF THE DRAWINGS

20 The file of this patent contains at least one color photograph. Copies of this patent with color photographs will be provided by the Patent and Trademark Office upon request and payment of necessary fee.

Figure 1 shows the amino acid sequence alignment of FATPs: mmFATP1 (SEQ ID NO:92), mmFATP2 (SEQ ID NO:93), mmFATP3 (SEQ ID NO:94),  
25 mmFATP4 (SEQ ID NO:95), mmFATP5 (SEQ ID NO:96), ceFATPa (SEQ ID NO:97), scFATP (SEQ ID NO:98) and mtFATP (SEQ ID NO:99). The underlining (amino acid residues 204-212 of mtFATP) indicates an AMP binding motif which is found in many classes of proteins; the underlining at amino acid residues 204-507 of the mtFATP sequence indicates the FATP 360 amino acid signature sequence.

-7-

Figures 2A-2D show results of LCFA uptake assays. Figures 2A-2D: COS cells were cotransfected using the DEAE-dextran method with the mammalian expression vectors pCDNA-CD2 either alone (control; Figure 2A) or in combination with one of the FATP-containing expression vectors (pCDNA-mmFATP1, Figure 2B; pCDNA-mmFATP2, Figure 2C; or pCMV-SPORT2-mmFATP5, Figure 2D) as described in Materials and Methods for Example 2. COS cells were gated on forward scatter (FSC) and side scatter (SS), and the results shown represent >10,000 cells. Cells exhibiting >300 CD2 fluorescence units (vertical line) representing 15% of all cells were deemed CD2 positive.

Figure 3 is a graph of fluorescence of cells expressing a FATP gene. As in Figures 2A-2D, COS cells were cotransfected with pCDNA-CD2 either alone (control) or in combination with one of the FATP-containing expression vectors (pCDNA-mmFATP1, pCDNA-mmFATP2, pCMV-SPORT2-mmFATP5, or pCDNA-ceFATPb). The mean BODIPY-FA fluorescence of the CD2-positive cells is plotted; results shown represent the average of three experiments, each consisting of greater than 50,000 COS cells. Note that a logarithmic scale is used on the ordinate.

Figure 4 is a graph of the uptake of palmitate with time. The full-length coding region of mtFATP (squares) or a control protein (TFE3; circles) was subcloned into the inducible, prokaryotic expression vector pET (Novagen, Madison, WI). Expression from the resulting plasmid was induced (solid symbols) in transformed *E. coli* cells with 1 mM isopropyl- $\beta$ -D-thiogalactoside (IPTG) for 1 hour, or cells were left uninduced (open symbols). Data points were done in triplicate and counts were normalized to the number of bacteria as determined by OD<sub>600</sub>.

Figure 5 is a phylogenetic tree produced by aligning complete and partial sequences for *FATP* genes from human, rat, mouse, puffer fish, *D. melanogaster*, *C. elegans*, *S. cerevisiae*, and *M. tuberculosis* using ClustalX and using these data to produce a phylogenetic tree using TreeViewPPC. The bar indicates the number of substitutions per residue, i.e., 0.1 corresponds to a distance of 10 substitutions per 100 residues.

Figure 6 shows a comparison of the FATP signature sequences of mmFATP1 (SEQ ID NO:1), mmFATP5, (SEQ ID NO:2), ceFATPa (SEQ ID NO:3), scFATP (SEQ ID NO:4) and mtFATP (SEQ ID NO:5).

Figure 7 shows the sequence identity among the FATP family members and VLACs, based on the 360 amino acid signature sequence of FATP from Figure 1.

Figures 8A and 8B are the mmFATP3 DNA sequence (SEQ ID NO:6).

Figure 9 is the mmFATP3 protein sequence (SEQ ID NO:7).

Figures 10A and 10B are the mmFATP4 DNA sequence (SEQ ID NO:8).

Figure 11 is the mmFATP4 protein sequence (SEQ ID NO:9).

Figures 12A and 12B are the mmFATP5 DNA sequence (SEQ ID NO:10).

Figure 13 is the mmFATP5 protein sequence (SEQ ID NO:11).

Figures 14A and 14B are the hsFATP2 DNA sequence (SEQ ID NO:12).

Figure 15 is the hsFATP2 protein sequence (SEQ ID NO:13).

Figures 16A and 16B are the hsFATP3 DNA sequence (SEQ ID NO:14).

Figure 17 is the hsFATP3 protein sequence (SEQ ID NO:15).

Figures 18A and 18B are the hsFATP4 DNA sequence (SEQ ID NO:16).

Figure 19 is the hsFATP4 protein sequence (SEQ ID NO:17).

Figures 20A and 20B are the hsFATP5 DNA sequence (SEQ ID NO:18).

Figure 21 is the hsFATP5 protein sequence (SEQ ID NO:19).

Figures 22A and 22B are the hsFATP6 DNA sequence (SEQ ID NO:20).

Figure 23 is the hsFATP6 protein sequence (SEQ ID NO:21).

Figures 24A and 24B are the mtFATP DNA sequence (SEQ ID NO:22).

Figure 25 is the mtFATP protein sequence (SEQ ID NO:23).

Figure 26 shows the DNA sequence (SEQ ID NO:24) and predicted amino acid sequence (SEQ ID NO:25) of human FATP1.

Figure 27 shows the DNA sequence (SEQ ID NO:26) and predicted amino acid sequence (SEQ ID NO:27) of human FATP4.

Figure 28A is a hydrophobicity plot for hsFATP1, showing that it has multiple membrane-spanning domains.

Figure 28B is the amino acid composition of hsFATP1.

Figure 28C is a hydrophilicity plot for hsFATP1, made using the Kyte-Doolittle method, averaging hydrophilicity values for 18 amino acid residues at a time.

Figure 29A is a hydrophobicity plot for hsFATP4, showing that it has  
5 multiple membrane-spanning domains.

Figure 29B is a listing of the amino acid composition of hsFATP4.

Figure 29C is a hydrophilicity plot for hsFATP4, made using the Kyte-Doolittle method, averaging hydrophilicity values for 18 amino acid residues at a time.

10 Figures 30A and 30B show a comparison of the nucleotide sequence of human FATP1 (SEQ ID NO:28) and the nucleotide sequence of mouse FATP1 (SEQ ID NO:29).

Figures 31A and 31B show a comparison of the nucleotide sequence of human FATP4 (SEQ ID NO:30) and the nucleotide sequence of mouse FATP4  
15 (SEQ ID NO:31).

Figure 32 shows a comparison of the amino acid sequence of human FATP1 (SEQ ID NO:32) and the amino acid sequence of mouse FATP1 (SEQ ID NO:33). Shaded amino acid residues match the consensus sequence exactly.

Figure 33 shows a comparison at the amino acid level of human FATP4  
20 (SEQ ID NO:34) and mouse FATP4 (SEQ ID NO:35). Shaded amino acid residues match the consensus sequence exactly.

Figure 34 shows the nucleotide sequence (SEQ ID NO:36) and predicted amino acid sequence (SEQ ID NO:37) of hsFATP6.

Figure 35A is a hydrophobicity plot for hsFATP6, showing that it has  
25 multiple membrane-spanning domains.

Figure 35B is a listing of the amino acid composition of hsFATP6.

Figure 35C is a hydrophilicity plot for hsFATP6, made using the Kyte-Doolittle method, averaging hydrophilicity values for 18 amino acid residues at a time.



Figure 36 shows an alignment of the amino acid sequences of hsFATP1 (SEQ ID NO:38), hsFATP4 (SEQ ID NO:39) and hsFATP6 (SEQ ID NO:40). Shaded amino acid residues match the consensus sequence exactly.

Figure 37 shows results of assessment of fatty acid uptake by human FATP1  
5 and human FATP4. The percent of CD2-positive cells exhibiting a BODIPY-fluorescence of more than 300 arbitrary units is plotted for the three different conditions tested.

Figure 38 is a graph showing uptake of tritiated oleate, with time, by 293 cells transfected with either (diamonds) a plasmid for expression of human FATP4  
10 or (squares) a control plasmid.

Figure 39 is an illustration of the amino acid sequences of human FATP4 (SEQ ID NO:41) and mouse FATP4 (SEQ ID NO:42) compared to human FATP1 (SEQ ID NO:43). Shown by underlining are the FATP consensus sequence (236-556 of hsFATP1) and the AMP-binding motif (246-254 of hsFATP1). The human  
15 FATPs were cloned by screening libraries with sequences from ESTs (expressed sequence tags). Mouse FATP4 was cloned by PCR using degenerate primers.

Figure 40 is a graph showing the uptake, with time, of tritiated oleate by mouse enterocytes in the presence of no oligonucleotide (squares), sense oligonucleotide (circles) or antisense oligonucleotide (diamonds).

Figure 41 is a bar graph showing uptake of tritiated oleate, by mouse  
20 enterocytes in the presence of various concentrations of antisense (solid bars), mismatch (stippled bars) or sense (lined bars) oligonucleotides.

Figure 42 is a bar graph showing uptake of tritiated oleate and uptake of <sup>35</sup>S-labeled methionine by mouse enterocytes to which were added no oligonucleotide,  
25 the antisense oligonucleotide, or the mismatch oligonucleotide.

Figure 43A is the nucleotide sequence of the gene encoding mouse FATP4 (SEQ ID NO:44).

Figure 43B is the amino acid sequence of mouse FATP4 protein (SEQ ID NO:45).

Figures 44A, 44B, and 44C are the hsFATP1 DNA sequence (SEQ ID  
30 NO:46). Coding region: 175-2115 (1941 nt).

Figure 45 is the hsFATP1 protein sequence (SEQ ID NO:47).

Figures 46A and 46B are the hsFATP2 DNA sequence (SEQ ID NO:48).

Coding region: 223-2085 (1863 nt).

Figure 47 is the hsFATP2 protein sequence (SEQ ID NO:49).

5 Figure 48 is the partial DNA sequence of hsFATP3 (SEQ ID NO:50).

Coding region: 1-993.

Figure 49 is the partial protein sequence of hsFATP3 (SEQ ID NO:51).

Figures 50A, 50B, and 50C are the hsFATP4 DNA sequence (SEQ ID NO:52). Coding region: 208-2139 (1932 nt).

10 Figure 51 is the hsFATP4 protein sequence (SEQ ID NO:53).

Figure 52 is the hsFATP5 partial DNA sequence (SEQ ID NO:54). Coding region: 1-1062.

Figure 53 is the hsFATP5 partial protein sequence (SEQ ID NO:55).

Figures 54A, 54B, and 54C are the hsFATP6 DNA sequence (SEQ ID NO:56). Coding region: 643-2502 (1860 nt).

Figure 55 is the hsFATP6 protein sequence (SEQ ID NO:57).

Figures 56A, 56B, and 56C are the mFATP1 DNA sequence (m=*Rattus norvegicus*; (SEQ ID NO:58). Coding region: 75-2015 (1941 nt).

Figure 57 is the mFATP1 protein sequence (SEQ ID NO:59).

20 Figures 58A, 58B, and 58C are the mFATP2 DNA sequence (SEQ ID NO:60). Coding region: 795-2657 (1863 nt).

Figure 59 is the mFATP2 protein sequence (SEQ ID NO:61).

Figures 60A and 60B are the mFATP4 partial DNA sequence (SEQ ID NO:62). Coding region: 1-1218.

25 Figure 61 is the mFATP4 partial DNA sequence (SEQ ID NO:63).

Figures 62A, 62B, and 62C are the mmFATP1 DNA sequence (SEQ ID NO:64). Coding region: 1-1944.

Figure 63 is the mmFATP1 protein sequence (SEQ ID NO:65).

Figures 64A and 64B are the mmFATP2 DNA sequence (SEQ ID NO:66).

30 Coding region: 121-1992 (1872 nt).

Figure 65 is the mmFATP2 protein sequence (SEQ ID NO:67).

Figures 66A and 66B are the mmFATP3 partial DNA sequence (SEQ ID NO:68). Coding region: 1-1830.

Figure 67 is the mmFATP3 partial protein sequence (SEQ ID NO:69).

Figures 68A, 68B, and 68C are the mmFATP4 DNA sequence (SEQ ID NO:70). Coding region: 1-1932.

Figures 69 is the mmFATP4 protein sequence (SEQ ID NO:71).

Figures 70A and 70B are the mmFATP5 DNA sequence (SEQ ID NO:72). Coding region: 60-2129.

Figure 71 is the mmFATP5 protein sequence (SEQ ID NO:73).

Figures 72A and 72B are the dmFATP partial DNA sequence (dm=*Drosophila melanogaster*; SEQ ID NO:74). Coding region: 1-1773.

Figure 73 is the dmFATP partial protein sequence (SEQ ID NO:75).

Figure 74 is the drFATP partial DNA sequence (dr=*Danio rerio*, zebrafish; SEQ ID NO:76) Coding region: 1-173.

Figure 75 is the drFATP partial protein sequence (SEQ ID NO:77).

Figure 76A and 76B are the ceFATPa DNA sequence (SEQ ID NO:78). Coding region: 1-1953.

Figure 77 is the ceFATPa protein sequence (SEQ ID NO:79).

Figures 78A and 78B are the ceFATPb DNA sequence (SEQ ID NO:80). Coding region: 1-1968.

Figure 79 is the ceFATPb protein sequence (SEQ ID NO:81).

Figures 80A and 80B are the chFATP DNA sequence (SEQ ID NO:82; ch=*Cochliobolus heterostrophus*). Coding region: 1-1932.

Figure 81 is the chFATP protein sequence (SEQ ID NO:83).

Figure 82 is the anFATP partial protein sequence (an=*Aspergillus nidulans*; SEQ ID NO:84). Coding region: 1-597.

Figure 83 is the anFATP partial protein sequence (SEQ ID NO:85).

Figure 84 is the mgFATP partial DNA sequence (mg=*Magnaporthe grisea*, rice blast; SEQ ID NO:86). Coding region: 1-522.

Figure 85 is the mgFATP partial protein sequence (SEQ ID NO:87).

Figures 86A and 86B are the scFATP DNA sequence (SEQ ID NO:88).  
Coding region: 1-1872.

Figure 87 is the scFATP protein sequence (SEQ ID NO:89).

Figures 88A and 88B are the mtFATP DNA sequence (SEQ ID NO:90).

5        Figure 89 is the mtFATP protein sequence (SEQ ID NO:91). Coding region:  
1-1794.

Figure 90 is a consensus sequence of the FATP signature sequence (SEQ ID NO:100), based on 23 independent sequences aligned in ClustalX. The height of the bar at each amino acid residue position indicates the degree of conservation at that  
10       position. Gaps have been inserted to maintain the strength of the alignment.

Figure 91 is a hydrophilicity plot for hsFATP2, made using the Kyte-Doolittle method, averaging hydrophilicity values for 18 amino acid residues at a time.

Figure 92 is a hydrophilicity plot for the hsFATP3 partial protein, made using the Kyte-Doolittle method, averaging hydrophilicity values for 18 amino acid  
15       residues at a time.

Figure 93 is a hydrophilicity plot for the hsFATP5 partial protein, made using the Kyte-Doolittle method, averaging hydrophilicity values for 18 amino acid residues at a time.

Figures 94A and 94B are a representation of the DNA sequence (SEQ ID NO:101) of the hsFATP3 gene, and the amino acid sequence (SEQ ID NO:102) of the  
20       hsFATP3 protein.

Figure 95 shows that mammalian expression constructs containing either hsFATP4 (squares and triangles) or empty control vector (circles) were stably transfected into 293 cells. Short-term uptake of Bodipy-FA in the presence of BSA  
25       was determined by FACS. The mean fluorescence of the viable cell population is expressed in arbitrary fluorescence units. FATP4 protein expression was determined by densitometry of anti-FATP4 Western blots, and is expressed in arbitrary units.

Figure 96 is a bar graph illustrating short-term uptake of Bodipy-palmitate (1  $\mu$ M), either by control cells (black bars) or FATP4-expressing cells (hatched bars),  
30       was measured in the presence of 0, 10, 100  $\mu$ M unlabeled palmitate. FA uptake was quantified by FACS and expressed in arbitrary fluorescence units.

Figure 97 shows the rate of [ $^2\text{H}$ ]palmitate uptake by 293 cells, which were stably transfected with a construct for either human FATP4 (diamonds) or an empty vector (circles), compared to that of isolated enterocytes (squares).

Figure 98 is a bar graph illustrating the results when isolated enterocytes were  
5 incubated for 48h with increasing concentrations of the FATP4 antisense oligonucleotide or with 100  $\mu\text{M}$  of a randomized control oligonucleotide with identical nucleotide composition to the FATP4 antisense oligonucleotide. The uptake of oleate by the enterocytes was then measured over a 5 min time interval (solid bars). In parallel, the levels of FATP4 protein and, as a loading control,  $\beta$ -catenin, were  
10 determined by Western blotting and quantitated using densitometry (hatched bars). FA uptake and FATP4 protein levels were normalized to that of untreated cells. The averages and standard deviations of 4 independent experiments are shown.

Figure 99 is a bar graph illustrating the uptake rates of [ $^3\text{H}$ ]oleate, [ $^3\text{H}$ ]palmitate and [ $^{35}\text{S}$ ]methionine by primary enterocytes were measured after 48h  
15 incubation with either 100  $\mu\text{M}$  FATP4 antisense (solid bars) or 100  $\mu\text{M}$  randomized control oligonucleotide (hatched bars) and expressed as % of untreated cells.

Figure 100 is a bar graph illustrating that 8 kb of FATP5 genomic sequence SEQ ID NO.: 106 is sufficient for liver specific transcription *in vitro*. A luciferase reporter construct containing 8 kb upstream of the FATP5 initiator methionine was  
20 transfected into various cell lines using calcium phosphate as described in Example 17. Forty-eight hours after transfection, luciferase activity was measured and normalized to  $\beta$ -galactosidase activity. For each cell line, fold induction was determined by dividing the relative luciferase activity of the 8 kb construct by that of the promoter-less luciferase reporter vector. The data shown represent the mean of  
25 three experiments done in triplicate. Error bars indicate the SEM.

Figure 101 is a bar graph illustrating deletion analysis of the FATP5 promoter. Constructs containing deletions of the FATP5 promoter were transfected into HepG2 cells, assayed for luciferase activity, and normalized to  $\beta$ -galactosidase (RLU). The labels on the vertical axis correspond to the length of the promoter segment as  
30 measured from the initiator methionine. The data shown represents the mean of three experiments done in triplicate. Error bars indicate the SEM.

Figure 102 is a bar graph illustrating that 271 base pairs upstream of the FATP5 initiator methionine are sufficient for liver specific luciferase activity. A luciferase reporter construct containing 271 base pairs upstream of the FATP5 initiator methionine was transfected into various cell lines using calcium phosphate as described in Methods Example 17. Forty eight hours after transfection, luciferase activity was measured and normalized to  $\beta$ -galactosidase activity. For each cell line, fold induction was determined by dividing the relative luciferase activity of the -271 base pair construct by that of the promoter-less luciferase reporter vector. The data shown represent the mean of three experiments done in triplicate. Error bars indicate the SEM.

Figures 103A and 103B illustrate mutations of the GC box which abolish transcriptional activity. A: Schematic of mutations in the GC box aligned with the normal sequence (SEQ ID NO.: 106, SEQ ID NO.: 107, SEQ ID NO.: 108). The GC box consensus sequence is underlined. B: Constructs containing 271 base pairs upstream of the FATP5 initiator methionine with the mutations in the GC box depicted in part A were transfected into HepG2 cells, assayed for luciferase activity, and normalized to  $\beta$ -galactosidase (RLU). The data shown represent the mean of three experiments done in triplicate. Error bars indicate the SEM.

Figure 104 shows a gel shift analysis of the GC box with HepG2 nuclear extracts. Schematic showing the sequence of the oligonucleotides used in gel shift studies. The numbering reflects the distance from the initiator methionine. The two pairs of oligonucleotides are indicated by the lines and labeled AF-1 (SEQ ID NO.: 111, SEQ ID NO.: 112) and AF-2 (SEQ ID NO.: 109, SEQ ID NO.: 110).

Figure 105 is a bar graph illustrating that 30bp internal deletions of the FATP5 promoter identify another region required for luciferase activity in HepG2 cells. Reporter constructs were transfected into HepG2 cells. Luciferase activity was measured and normalized to  $\beta$ -galactosidase activity (RLU). The labels on the horizontal axis correspond to the nucleotides that were deleted and the numbering on the vertical axis represents the distance from the initiator methionine. The data shown represent the mean of three experiments done in triplicate. Error bars indicate

the SEM. Note that the five fold higher RLU activity in this figure relative to Figures 101 and 103 is the result of a manufacturer change in the  $\beta$ -galactosidase reagent.

Figure 106 is a bar graph illustrating that a linker scan of the FATP5 promoter identifies two additional elements required for transcription in HepG2 cells. Reporter  
5 constructs were transfected into HepG2 cells. Luciferase activity was measured and normalized to  $\beta$ -galactosidase activity (RLU). The labels on the horizontal axis correspond to the constructs in part A. The data shown represent the mean of three experiments done in triplicate. Error bars indicate the SEM. Please note that the lower RLU activity in this figure relative to Figures 101 and 103 is also the result of a  
10 manufacturer change in the  $\beta$ -galactosidase reagent.

Figure 107 is a schematic of the FATP5 promoter (SEQ ID NO.: 113). The GC box and two motifs identified in the linker scan are boxed and labeled. An arrow indicates the translational initiator of the FATP5 protein. The two halves of the palindrome contained in the novel motifs and referred to in the discussion are  
15 underlined.

Figure 108 is a photograph showing FATP2 expression in the mouse gall bladder epithelium.

Figure 109 is a photograph showing FATP2 expression in chimpanzee liver.

Figure 110 is a photograph showing FATP5 expression in chimpanzee liver.

20 Figures 111A and 111B represent the DNA sequence (SEQ ID NO:116) of human FATP3.

Figure 112 represents the amino acid sequence (SEQ ID NO:117) of human FATP3.

Figure 113 is a bar graph showing the results of an experiment comparing  
25 fatty acid transport between cells transfected with SEQ ID NO: 116 and untransfected cells.

Figures 114A, 114B, 114C and 114D represent portions of the amino acid sequence of mmFATP4 which were produced as fusion polypeptides in *E. coli* cells.

Figure 115 is a schematic illustrating certain components of the fusion  
30 polypeptides depicted in Figures 114A-D. The schematic shows the lipocalin domain

-17-

as well as other identified motifs and notes the relative location of each in the mmFATP4 fusion polypeptide.

Figure 116 is a bar graph illustrating the results of an experiment comparing the binding capabilities of the fusion polypeptides shown in Figures 114A-D for an  
5 oleate fatty acid.

Figure 117 is a bar graph showing the results of an experiment comparing binding of various fatty acids between two of the fusion polypeptides depicted in Figure 114A-D.

Figure 118A-G illustrates the consensus sequence of hsFATP1, hsFATP2,  
10 hsFATP3, hsFATP4, hsFATP5 and hsFATP6 with the lipocalin domain and AMP-binding domain of each noted.

The foregoing and other objects, features and advantages of the invention will be apparent from the following more particular description of preferred embodiments of the invention, as illustrated in the accompanying drawings in which like reference  
15 characters refer to the same parts throughout the different views. The drawings are not necessarily to scale, emphasis instead being placed upon illustrating the principles of the invention.

#### DETAILED DESCRIPTION OF THE INVENTION

As described herein, FATPs are a large evolutionarily conserved family of  
20 proteins that mediate the transport of LCFAs into cells. The family includes proteins which are conserved from mycobacteria to humans and exhibit very different expression patterns in tissues. Specific embodiments described include FATPs from mice, humans, nematodes, fungi and mycobacteria which have been shown to be functional LCFA transporters. The term "fatty acid transport proteins" ("FATPs") as  
25 used herein, refers to the proteins described herein as FATP1, FATP2, FATP3, FATP4, FATP5 and FATP6, which have been described in one or more species of mammals, as well as mtFATP, ceFATP, scFATP, anFATP, mgFATP, and chFATP, and other proteins sharing at least about 50% amino acid sequence similarity,



preferably at least about 60% sequence similarity, more preferably at least about 70% sequence similarity, and still more preferably, at least about 80% sequence similarity, and most preferably, at least about 90% sequence similarity in the approximately 360 amino acid signature sequence. The approximately 360 amino acid FATP signature sequence is shown in Figure 1. The consensus sequence of the signature sequence is shown in Figure 90. The nomenclature used herein to refer to FATPs includes a species-specific prefix (e.g., mm, *Mus musculus*; hs or h, *Homo sapiens* or human; mt *M. tuberculosis*; dm, *D. melanogaster*; ce, *C. elegans*; sc, *Saccharomyces cerevisiae*) and a number such that mammalian homologues in different species share the same number. For example, six human and five mouse *FATP* genes which are expressed in a variety of tissues are described herein and are referred to, respectively, as hsFATP1-hsFATP6 and mmFATP1-mmFATP5; for example, hsFATP4 and mmFATP4 are the human and mouse orthologs.

Expression patterns of human and mouse FATPs have been assessed and are described below. Briefly, results of these assessments show that FATP5 is a liver-specific gene. FATP2 is highly expressed in liver, kidney and gall bladder epithelium. Both of these proteins, as well as FATP4 and FATPs from nematodes and mycobacteria, have been shown to be functional LCFA transporters. Results have also shown that FATP4 mRNA is present at high levels in epithelial cells of two regions of the small intestine (the jejunum and ileum) and at lower, but significant, levels in a third region (the duodenum). They further showed that FATP2 mRNA is present in epithelial cells of the duodenum at a level similar to that of FATP4 mRNA levels, but is present at lower levels in the jejunum and ileum. FATP4 mRNA was absent from other cell types of the small intestine and no FATP4 mRNA could be detected in any cells of the colon. No signals above background could be detected for FATP1, FATP3 and FATP5 in any of the intestinal tissues. Thus, FATP4 is the major FATP in the mouse small intestine, which supports a major role for FATP4 (along with FATP2 to a lesser extent) in absorption of free fatty acids. hsFATP4 was clearly expressed in the jejunum and ileum; expression was absent in the stomach. This, too, is consistent with a major role for FATP4 in absorption of fatty acids in the human gut. Analysis of FATP expression in human tissues, also described in detail

below, showed that hsFATP6, which has no mouse ortholog as yet, is expressed at high levels in the heart and at low levels in the placenta, but is undetectable in the other tissues assessed (Example 9). This is consistent with a major role for FATP6 in absorption of fatty acids in the heart.

5           Analysis of FATP3 expression in murine tissues, also described in detail below, showed that expression occurs at detectable levels in liver, spleen, heart, kidney, testis, white adipose tissue, exocrine and endocrine pancreatic cells, and also in lung tissues. FATP3 is expressed at high levels in type-II pneumocytes, a cell type noted for secretion a surfactant, a phospholipid-rich film critical for lung function  
10 (Example 19).

Long chain fatty acids (LCFAs) are an important energy source for pro- and eukaryotes and are involved in diverse cellular processes, such as membrane synthesis, intracellular signaling, protein modification, and transcriptional regulation. In developed Western countries, human dietary lipids are mainly di- and triglycerides  
15 and account for approximately 40% of caloric intake (Weisburger, J. H. (1997) *J. Am. Diet. Assoc.* 97:S16-S23). These lipids are broken down into fatty acids and glycerol by pancreatic lipases in the small intestine (Chapus, C., Rovey, M., Sarda, L & Verger, R. (1988) *Biochimie* 70:1223-34); LCFAs are then transported into brush border cells, where the majority is re-esterified and secreted into the lymphatic system  
20 as chylomicrons (Green, P.H. & Riley, J.W. (1981) *Aust. N.Z.J. Med.* 11:84-90). Fatty acids are liberated from lipoproteins by the enzyme lipoprotein lipase, which is bound to the luminal side of endothelial cells (Scow, R.O. & Blachette-Mackie, E.J. (1992) *Mol. Cell. Biochem* 116:181-191). "Free" fatty acids in the circulation are bound to serum albumin (Spector, A.A. (1984) *Clin. Physiol. Biochem* 2:123-134)  
25 and are rapidly incorporated by adipocytes, hepatocytes, and cardiac muscle cells. The latter derive 60-90% of their energy through the oxidation of LCFAs (Neely, J.F. Rovetto, M.J. & Oram, J.F. (1972) *Prog. Cardiovasc. Dis.* 15:289-329). Although saturable and specific uptake of LCFAs has been demonstrated for intestinal cells, hepatocytes, cardiac myocytes, and adipocytes, the molecular mechanisms of LCFA  
30 transport across the plasma membrane have remained controversial (Hui, T.Y. & Bernlohr, D.A. (1997) *Front. Biosci.* 15:d222-31-d231; Schaffer, J.E. & Lodish, H.F,

(1995) *Trends Cardiovasc. Med.* 5:218-224). Described herein is a large family of highly homologous mammalian LCFA transporters which show wide expression, including in all tissues relevant to fatty acid metabolism. Further described are novel members of this family in other species, including mycobacterial and nematode  
5 FATPs which, like their mammalian counterparts, are functional fatty acid transporters.

The discovery of a diverse but highly homologous family of FATPs is reminiscent of the glucose transporter family. In a manner similar to the FATPs, the glucose transporters have very divergent patterns of tissue expression (McGowan,  
10 K.M., Long, S.D. & Pekala, P.H. (1995) *Pharmacol. Ther.* 66:465-505). The FATPs, like glucose transporters, may also differ in their substrate specificities, uptake kinetics, and hormonal regulation (Thorens, B. (1996) *Am. J. Physiol.* 270:G541-G553). Indeed, the levels of fatty acids in the blood, like those of glucose, can be regulated by insulin and are dysregulated in diseases such as noninsulin-dependent  
15 diabetes and obesity (Boden, G. (1997) *Diabetes* 46:3-10). The underlying mechanisms for the regulation of free fatty acid concentrations in the blood are not understood, but could be explained by hormonal modulation of FATPs.

Insulin-resistance is thought to be the major defect in non insulin-dependent diabetes mellitus (NIDDM) and is one of the earliest manifestations of NIDDM  
20 (McGarry (1992) *Science* 258:766-770). Free fatty acids (FFAs) may provide an explanation for why obesity is a risk factor for NIDDM. Plasma levels of FFAs are elevated in diabetic patients (Reaven *et al.* (1988) *Diabetes* 37:1020). Elevated plasma free fatty acids (FFAs) have been demonstrated to induce insulin-resistance in whole animals and humans (Boden (1998) *Front. Biosci.* 3:D169-D175). This  
25 insulin-resistance is likely mediated by effects of FFAs on a variety of issues. FFAs added to adipocytes *in vitro* induce insulin resistance in this cell type as evidenced by inhibition of insulin-induced glucose transport (Van Epps-Fung *et al.* (1997) *Endocrinology* 138:4338-4345). Rats fed a high fat diet developed skeletal muscle insulin resistance as evidenced by a decrease in insulin-induced glucose uptake by  
30 skeletal muscle (Han *et al.*, (1997) *Diabetes* 46:1761-1767). In addition, elevated plasma FFAs increase insulin-suppressed endogenous glucose production in the liver

(Boden (1998) *Front. Biosci.* 3:D169-D175), thus increasing hepatic glucose output. It has been postulated that the adverse effects of plasma free fatty acids are due to the FFAs being taken up into the cell, leading to an increase in intracellular long chain fatty acyl CoA; intracellular long chain acyl CoAs are thought to mediate the effects of FFAs inside the cell. Thus, fatty acid induced insulin-resistance may be prevented by blocking uptake of FFAs into select tissues, in particular liver (by blocking FATP2 and/or FATP5), adipocyte (by blocking FATP1), and skeletal muscle (by blocking FATP1). Blocking intestinal fat absorption (by blocking FATP4) is also expected to reduce plasma FFA levels and thus improve insulin resistance.

During the pathogenesis of NIDDM insulin-resistance can initially be counteracted by increasing insulin output by the pancreatic beta cell. Ultimately, this compensation fails, beta cell function decreases and overt diabetes results (McGarry (1992) *Science* 258: 766-770). Manipulating beta cell function is a second point where fatty acid transporter blockers may be beneficial for diabetes. While no FATP homolog has been identified so far that is expressed in the beta cell of the pancreas, the data described below suggest the existence of such a transporter and the sequence information included herein provides the means to identify such a transporter by degenerate PCR, using primers to regions conserved in all FATP family members or by low stringency hybridization. It has been demonstrated that exposure of pancreatic beta-cells to FFAs increases the basal rate of insulin secretion; this in turn leads to a decrease in the intracellular stores of insulin, resulting in decreased capacity for insulin secretion after chronic exposure (Bollheimer *et al.*, (1998) *J. Clin. Invest.* 101:1094-1101). The effects of FFAs are again likely to be mediated by intracellular long chain fatty acyl CoA molecules (Liu *et al.*, (1998) *J. Clin. Invest.* 101:1870-1875). FFAs have also been demonstrated to increase beta cell apoptosis (Shimabukuro *et al.*, (1998) *Proc. Nat. Acad. Sci. USA* 95:2498-2502), possibly contributing to the decrease in beta cell numbers in late stage NIDDM.

Another finding with potentially broad implications is the identification of a FATP homologue in *M. tuberculosis*. Tuberculosis causes more deaths worldwide than any other infectious agent and drug-resistant tuberculosis is re-emerging as a problem in industrialized nations (Bloom, B.R. & Small, P.M. (1998) *N. Engl. J.*

*Med.* 338:677-678). *Mycobacterium tuberculosis* has about 250 enzymes involved in fatty acid metabolism, compared with only about 50 in *E. coli*. It has been suggested that, living as a pathogen, the mycobacteria are largely lipolytic, rather than lipogenic, relying on the lipids within mammalian cells and the tubercle (Cole, S.T. *et al.*,  
5 *Nature* 393:537-544 (1998)). The *de novo* synthesis of fatty acids in *Mycobacterium leprae* is insufficient to maintain growth (Wheeler, P.R., Bulmer, K & Ratledge, C. (1990) *J. Gene. Microbiol.* 136:211-217). Thus, it is reasonable to expect that inhibitors of mtFATP will serve as therapeutics for tuberculosis. FATPs expressed in mycobacteria can be targeted to reduce or prevent replication of mycobacteria (e.g., to  
10 reduce or prevent replication of *M. tuberculosis*) and, thus, reduce or prevent their adverse effects. For example, a FATP or FATPs expressed by *M. tuberculosis* can be targeted and inhibited, thus reducing or preventing growth of this pathogen (and tuberculosis in humans and other mammals). An inhibitor of an *M. tuberculosis* FATP can be identified, using methods described herein (e.g., expressing the FATP  
15 in an appropriate host cell, such as *E. coli* or COS cells; contacting the cells with an agent or drug to be assessed for its ability to inhibit the FATP and, as a result, mycobacterial growth, and assessing its effects on growth). A drug or agent identified in this manner can be further tested for its ability to inhibit a *M. tuberculosis* FATP and *M. tuberculosis* infection in an appropriate animal model or  
20 in humans. A method of inhibiting mycobacterial growth, particularly growth of *M. tuberculosis*, and compounds useful as drugs for doing so are also the subject of this invention.

An isolated polynucleotide encoding mtFATP, like other polynucleotides encoding FATPs of the FATP family, can be incorporated into vectors, nucleic acids  
25 of viruses, and other nucleic acid constructs that can be used in various types of host cells to produce mtFATP. This mtFATP can be used, as it appears on the surface of cells, or in various artificial membrane systems, to assess fatty acid transport function, to identify ligands and molecules that are modulators of fatty acid transport activity. Molecules found to be inhibitors of mtFATP function can be incorporated  
30 into pharmaceutical compositions to administer to a human for the treatment of tuberculosis.

Particular embodiments of the invention are polynucleotides encoding a FATP of *Cochliobolus (Helminthosporium) heterostrophus* or portions or variants thereof, the isolated or recombinantly produced FATP, methods for assessing whether an agent binds to the chFATP, and further methods for assessing the effect of an agent being tested for its ability to modulate fatty acid transport activity.

*Cochliobolus heterostrophus* is an ascomycete that is the cause of southern corn leaf blight, an economically important threat to the corn crop in the United States. The related species *C. sativus* causes crown rot and common root rot in wheat and barley. One or more FATPs of *C. heterostrophus* can be targeted for the identification of an inhibitor of chFATP function, which can be then be used as an agent effective against infection of plants by *C. heterostrophus* and related organisms. Methods described herein that were applied in studying the expression of a FATP gene and the function of the FATP in its natural site of expression or in a host cell, can be used in the study of the chFATP gene and protein.

*Magnaporthe grisea* (rice blast) is an economically important fungal pathogen of rice. Further embodiments of the invention are nucleic acid molecules encoding a FATP of *Magnaporthe grisea*, portions thereof, or variants thereof, isolated mgFATP, nucleic acid constructs, and engineered cells expressing mgFATP. Other aspects of the invention are assays to identify an agent which binds to mgFATP and assays to identify an agent which modulates the function of mgFATP in cells in which mgFATP is expressed or in artificial membrane systems. Agents identified as inhibiting mgFATP activity can be developed into anti-fungal agents to be used to treat rice infected with rice blast.

*Caenorhabditis elegans* is a nematode related to plant pathogens and human parasites. An isolated polynucleotide which encodes ceFATP, like other polynucleotides encoding FATPs of the FATP family described herein, can be incorporated into nucleic acid vectors and other constructs that can be used in various types of cells to produce ceFATP. ceFATP as it occurs in cells or as it can be isolated or incorporated into various artificial or reconstructed membrane systems, can be used to assess fatty acid transport, and to identify ligands and agents that modulate fatty acid transport activity. Agents found by such assays to be inhibitors of ceFATP

activity can be incorporated into compositions for the treatment of diseases caused by genetically related organisms with a FATP of similar sensitivity to the agents.

*Aspergillus nidulans* is one of a family of fungal species that can infect humans. Further embodiments of the invention of the family of polynucleotides  
5 encoding FATPs are polynucleotides encoding a FATP of *Aspergillus nidulans*, and vectors and host cells that can be constructed to comprise such polynucleotides. Further embodiments are a polypeptide encoded by such polynucleotides, portions thereof having one or more functions characteristic of a FATP, and various methods. The methods include those for identifying agents that bind to a FATP and those for  
10 assessing the effect of an agent being tested for its ability to modulate fatty acid transport activity. Those agents found to inhibit fatty acid transport function can be used in compositions as anti-fungal pharmaceuticals, or can be modified for greater effectiveness as a pharmaceutical.

One aspect of the invention relates to isolated nucleic acids that encode a  
15 FATP as described herein, such as those FATPs having an amino acid sequence in Figure 45 (SEQ ID NO:47), Figure 47 (SEQ ID NO:49), Figure 112 (SEQ ID NO: 117), Figure 51 (SEQ ID NO:53), Figure 53 (SEQ ID NO:55), and Figure 55 (SEQ ID NO:57) and nucleic acids closely related thereto as described herein.

Using the information provided herein, such as a nucleic acid sequence set  
20 forth in Figures 44A-44C (SEQ ID NO:46), Figures 46A and 46B (SEQ ID NO:48), Figure 112 (SEQ ID NO:116), Figures 50A-50C (SEQ ID NO:52), Figure 52 (SEQ ID NO:54), and Figures 54A-54C (SEQ ID NO:56), a nucleic acid of the invention encoding a FATP polypeptide has been obtained using standard cloning and screening methods, such as those for cloning and sequencing cDNA library  
25 fragments, followed by obtaining a full length clone. For example, to obtain a nucleic acid of the invention, a library of clones of cDNA of human or other mammalian DNA can be probed with a labeled oligonucleotide, such as a radiolabeled oligonucleotide, preferably about 17 nucleotides or longer, derived from a partial sequence. Clones carrying DNA identical to that of the probe can then be  
30 distinguished using stringent (also, "high stringency") hybridization conditions. By sequencing the individual clones thus identified with sequencing primers designed

from the original sequence it is then possible to extend the sequence in both directions to determine the full length sequence. Suitable techniques are described, for example, in *Current Protocols in Molecular Biology* (F.M. Ausubel et al, eds), containing supplements through Supplement 42, 1998, John Wiley and Sons, Inc.,  
5 especially chapters 5, 6 and 7.

Embodiments of the invention include isolated nucleic acid molecules comprising any of the following nucleotide sequences: 1.) a nucleotide sequence which encodes a protein comprising the amino acid sequence of hsFATP1 (SEQ ID NO:47), the amino acid sequence of hsFATP2 (SEQ ID NO:49), the amino acid  
10 sequence of hsFATP3 (SEQ ID NO:117), the amino acid sequence of hsFATP4 (SEQ ID NO: 53), the amino acid sequence of hsFATP5 (SEQ ID NO:55) or the amino acid sequence of hsFATP6 (SEQ ID NO:57); 2.) nucleotide sequences of hsFATP1, hsFATP2, hsFATP3, hsFATP4, hsFATP5, or hsFATP6 (SEQ ID NO:46, 48, 116, 52, 54, or 56, respectively); 3.) a nucleotide sequence which is complementary to the  
15 nucleotide sequence of hsFATP1 (SEQ ID NO:46), hsFATP2 (SEQ ID NO:48), hsFATP3 (SEQ ID NO:116), hsFATP4 (SEQ ID NO:52), hsFATP5 (SEQ ID NO:54) or hsFATP6 (SEQ ID NO:56); 4.) a nucleotide sequence which consists of the coding region of hsFATP1 (SEQ ID NO:46), the coding region of hsFATP2 (SEQ ID NO:48), the coding region of hsFATP3 (SEQ ID NO:116), the coding region of  
20 hsFATP4 (SEQ ID NO:52), the coding region of hsFATP5 (SEQ ID NO:54), or the coding region of hsFATP6 (SEQ ID NO:56).

The invention further relates to nucleic acids (nucleic acid molecules or polynucleotides) having nucleotide sequences identical over their entire length to those shown in the figures, for instance Figures 44A-44C (SEQ ID NO:46), Figures  
25 46A and 46B (SEQ ID NO:48), Figures 111A-B (SEQ ID NO:116), Figures 50A-50C (SEQ ID NO:52), Figure 52 (SEQ ID NO:54), and Figures 54A-54C (SEQ ID NO:56). It further relates to DNA, which due to the degeneracy of the genetic code, encodes a FATP encoded by one of the FATP-encoding DNAs, whose amino acid sequence is provided herein. Also provided by the invention are nucleic acids having  
30 the coding sequences for the mature polypeptides or fragments in reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-,



or pro- or prepro- protein sequence. The nucleic acids of the invention encompass nucleic acids that include a single continuous region or discontinuous regions encoding the polypeptide, together with additional regions, that may also contain coding or non-coding sequences. The nucleic acids may also contain non-coding sequences, including, for example, but not limited to, non-coding 5' and 3' sequences, such as the transcribed, non-translated sequences, termination signals, ribosome binding sites, sequences that stabilize mRNA, introns, polyadenylation signals, and additional coding sequences which encode additional amino acids. For example, a marker sequence that facilitates purification of the fused polypeptide can be encoded.

10 In certain embodiments of the invention, the marker sequence can be a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc., Venlo, The Netherlands) and described in Gentz *et al.*, *Proc. Natl. Acad. Sci. USA* 86: 821-824 (1989), or an HA tag (Wilson *et al.*, *Cell* 37: 767 (1984)), or a sequence encoding glutathione S-transferase of *Schistosoma japonicum* (vectors available from Pharmacia; see Smith, D.B. and Johnson K.S., *Gene* 67:31 (1988) and Kaelin, W.G. *et al.*, *Cell* 70:351 (1992)). Nucleic acids of the invention also include, but are not limited to, nucleic acids comprising a structural gene and its naturally associated sequences that control gene expression.

The invention further relates to nucleic acids (nucleic acid molecules or polynucleotides) that encode a FATP polypeptide. In a particular embodiment, a nucleic acid encodes a portion of a FATP which includes a motif or domain, for example, a lipocalin domain or an AMP-binding domain. Such a polypeptide portion can be a functional portion of a FATP protein. The term "lipocalin domain" is an art recognized term and as used herein refers to a particular domain present in FATP proteins. This domain is described as including regions of sequence homology as well as a common tertiary structure represented as an eight stranded antiparallel beta-barrel. (see Banaszak, L. *et al.*, *Advances in Protein Chemistry*, 45: 89-151). Many lipocalin domains can be identified structurally as a sequence contained within the general formula: [DENG]-X-[DENQGSTARK]-X(0,2)-[DENQARK]-[LIVFY]-

25 {CP}-G-{C}-W-[FYWLRH-X]-[LIVMTA], *e.g.*, the lipocalin signature sequence or consensus pattern (SEQ ID NO: 125). One skilled in the art will recognize that a

30

-27-

lipocalin domain for a particular FATP protein can vary in sequence from this general formula. A FATP lipocalin domain can be, for example, identical to the lipocalin signature sequence or can exhibit 60, 65, 70, 75, 80, 85, 90, 95 or greater per cent sequence identity compared to the general formula provided that it retains lipocalin binding function. For example, a lipocalin domain for each of the human FATPs, hsFATP1 (SEQ ID NO: 126), hsFATP2 (SEQ ID NO: 127), hsFATP3 (SEQ ID NO: 128), hsFATP4 (SEQ ID NO: 129), hsFATP5 (SEQ ID NO: 130), and hsFATP6 (SEQ ID NO: 131) has been identified. The pattern of these lipocalin domains are highly conserved across the FATP family.

10 A nucleic acid encoding a portion of a FATP polypeptide can encode one or more domains, and also can include additional nucleotides. For example, the nucleic acid can also include nucleotide sequences that encode a portion of a FATP protein that is upstream from a lipocalin domain. As the term "upstream" or "upstream sequences" is used herein in relation to the lipocalin domain, it is intended to refer to the nucleotide sequence which encodes all or a portion of a FATP protein located between the signal peptide (when one is present) and the lipocalin domain. In the absence of a signal peptide, the term refers to the nucleotide sequence which encodes all or a portion of a FATP protein between the lipocalin domain and the amino terminus (see Figure 115).

20 The invention further relates to variants, including naturally-occurring allelic variants, of those nucleic acids described specifically herein by DNA sequence, that encode variants of such polypeptides as those having the amino acid sequences shown in Figure 45 (SEQ ID NO:47), Figure 47 (SEQ ID NO:49), Figure 112 (SEQ ID NO:117), Figure 51 (SEQ ID NO:53) Figure 53 (SEQ ID NO:55), or Figure 55 (SEQ ID NO:57). Such variants include nucleic acids encoding variants of the above-listed amino acid sequences, wherein those variants have several, such as 5 to 10, 1 to 5, or 3, 2 or 1 amino acids substituted, deleted, or added, in any combination. Variants include polynucleotides encoding polypeptides with at least 95% but less than 100% amino acid sequence identity to the polypeptides described herein by amino acid sequence. Variant polynucleotides hybridize, under low to high stringency conditions, to the alleles described herein by DNA sequence. In one

30

embodiment, variants have silent substitutions, additions and deletions that do not alter the properties and activities of the FATP. Allelic variants of the polynucleotides encoding hsFATP1 (Figure 45; SEQ ID NO:47), hsFATP2 (Figure 47; SEQ ID NO:49), hsFATP3 (Figure 112; SEQ ID NO:117), hsFATP4 (Figure 51; SEQ ID NO:53), hsFATP5 (Figure 53; SEQ ID NO:55) and hsFATP6 (Figure 55; SEQ ID NO:57) will be identified as mapping to chromosomal locations listed for the corresponding wild type genes in Table 2 in Example 1.

Orthologous genes are gene loci in different species that are sufficiently similar to each other in their nucleotide sequences to suggest that they originated from a common ancestral gene. Orthologous genes arise when a lineage splits into two species, rather than when a gene is duplicated within a genome. Proteins that are orthologs are encoded by genes of two different species, wherein the genes are said to be orthologous.

The invention further relates to polynucleotides encoding polypeptides which are orthologous to those polypeptides having a specific amino acid sequence described herein, such as the amino acid sequences shown in Figure 45 (SEQ ID NO:47), Figure 47 (SEQ ID NO:49), Figure 112 (SEQ ID NO:117), Figure 51 (SEQ ID NO:53), Figure 53 (SEQ ID NO:55), or Figure 55 (SEQ ID NO:57). These polynucleotides, which can be called ortholog polynucleotides, encode orthologous polypeptides that can range in amino acid sequence identity to a reference amino acid sequence described herein, from about 65% to less than 100%, but preferably 70% to 80%, more preferably 80% to 90%, and still more preferably 90% to less than 100%. Orthologous polypeptides can also be those polypeptides that range in amino acid sequence similarity to a reference amino acid sequence described herein from about 75% to 100%, within the signature sequence. The amino acid sequence similarity between the signature sequences of orthologous polypeptides is preferably 80%, more preferably 90%, and still more preferably, 95%. The ortholog polynucleotides encode polypeptides that have similar functional characteristics (e.g., fatty acid transport activity) and similar tissue distribution, as appropriate to the organism from which the ortholog polynucleotides can be isolated.

Ortholog polynucleotides can be isolated from (e.g., by cloning or nucleic acid amplification methods) a great number of species, as shown by the sample of FATPs from evolutionarily divergent species described herein (see, e.g., Figures 44A-C through Figure 89). Ortholog polynucleotides corresponding to those in Figure 45  
5 (SEQ ID NO:47), Figure 47 (SEQ ID NO:49), Figures 111A-B (SEQ ID NO:116), Figure 51 (SEQ ID NO:53), Figure 52 (SEQ ID NO:55) and Figure 55 (SEQ ID NO:57) are those which can be isolated from mammals such as rat, dog, chimpanzee, monkey, baboon, pig, rabbit and guinea pig, for example.

Further variants that are fragments of the nucleic acids of the invention may  
10 be used to synthesize full-length nucleic acids of the invention, such as by use as primers in a polymerase chain reaction. As used herein, the term primer refers to a single-stranded oligonucleotide which acts as a point of initiation of template-directed DNA synthesis under appropriate conditions (e.g., in the presence of four different nucleoside triphosphates and an agent for polymerization, such as DNA or  
15 RNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. The appropriate length of a primer depends on the intended use of the primer, but typically ranges from 15 to 30 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template, but  
20 must be sufficiently complementary to hybridize with a template. The term primer site refers to the area of the target DNA to which a primer hybridizes. The term primer pair refers to a set of primers including a 5' (upstream) primer that hybridizes with the 5' end of the DNA sequence to be amplified and a 3' (downstream) primer that hybridizes with the complement of the 3' end of the sequence to be amplified.

25 Further embodiments of the invention are nucleic acids that are at least 80% identical over their entire length to a nucleic acid described herein, for example a nucleic acid having the nucleotide sequence in Figures 44A-44C (SEQ ID NO:46), Figures 46A-46B (SEQ ID NO:48), Figures 111A-B (SEQ ID NO:116), Figures 50A-50C (SEQ ID NO:52), Figure 52 (SEQ ID NO:54), and Figures 54A-54C (SEQ ID  
30 NO:56). Additional embodiments are nucleic acids, and the complements of such nucleic acids, having at least 90% nucleotide sequence identity to the above-

described sequences, and nucleic acids having at least 95% nucleotide sequence identity. In preferred embodiments, DNA of the present invention has 97% nucleotide sequence identity, 98% nucleotide sequence identity, or at least 99% nucleotide sequence identity with the DNA whose sequences are presented herein.

5 Other embodiments of the invention are nucleic acids that are at least 80% identical in nucleotide sequence to a nucleic acid encoding a polypeptide having an amino acid sequence as set forth in Figure 45 (SEQ ID NO:47), Figure 47 (SEQ ID NO:49), Figure 112 (SEQ ID NO:117), Figure 51 (SEQ ID NO:53), Figure 53 (SEQ ID NO:55) or Figure 55 (SEQ ID NO:57), or as such amino acid sequences are set  
10 forth elsewhere herein, and nucleic acids that are complementary to such nucleic acids. Specific embodiments are nucleic acids having at least 90% nucleotide sequence identity to a nucleic acid encoding a polypeptide having an amino acid sequence as described in the list above, nucleic acids having at least 95% sequence identity, and nucleic acids having at least 97% sequence identity.

15 The terms "complementary" or "complementarity" as used herein, refer to the natural binding of polynucleotides under permissive salt and temperature conditions by base-pairing. Complementarity between two single-stranded molecules may be "partial" in which only some of the nucleic acids bind, or it may be complete when total complementarity exists between the single-stranded molecules (that is, when A-  
20 T and G-C base pairing is 100% complete). The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, which depend on binding between nucleic acid strands.

The invention further includes nucleic acids that hybridize to the above-  
25 described nucleic acids, especially those nucleic acids that hybridize under stringent hybridization conditions. "Stringent hybridization conditions" or "high stringency conditions" generally occur within a range from about  $T_m$  minus 5°C (5° C below the strand dissociation temperature or melting temperature ( $T_m$ ) of the probe nucleic acid molecule) to about 20° C to 25° C below  $T_m$ . As will be understood by those of skill  
30 in the art, the stringency of hybridization may be altered in order to identify or detect molecules having identical or related polynucleotide sequences. An example of high

stringency hybridization follows. Hybridization solution is (6x SSC/10 mM EDTA/0.5% SDS/5x Denhardt's solution/100 µg/ml sheared and denatured salmon sperm DNA). Hybridization is at 64-65°C for 16 hours. The hybridized blot is washed two times with 2x SSC/0.5% SDS solution at room temperature for 15 minutes each, and two times with 0.2x SSC/0.5% SDS at 65°C, for one hour each. Further examples of high stringency conditions can be found on pages 2.10.1-2.10.16 (see particularly 2.10.8-11) and pages 6.3.1-6 in *Current Protocols in Molecular Biology* (Ausubel, F.M. *et al.*, eds., containing supplements up through Supplement 42, 1998). Examples of high, medium, and low stringency conditions can be found on pages 36 and 37 of WO 98/40404, which are incorporated herein by reference.

The invention further relates to nucleic acids obtainable by screening an appropriate library with a probe having a nucleotide sequence such as that set forth in Figures 44A-44C (SEQ ID NO:46), Figures 46A-46B (SEQ ID NO:48), Figure 111 (SEQ ID NO:116), Figures 50A-50C (SEQ ID NO:52), Figure 52 (SEQ ID NO:54) or Figures 54A-54C (SEQ ID NO:56), or a probe which is a sufficiently long fragment of any of the above; and isolating the nucleic acid. Such probes generally can comprise at least 15 nucleotides. Nucleic acids obtainable by such screenings may include RNAs, cDNAs and genomic DNA, for example, encoding FATPs of the FATP family described herein.

Further uses for the nucleic acid molecules of the invention, whether encoding a full-length FATP or whether comprising a contiguous portion of a nucleic acid molecule such as one given in SEQ ID NO:46, 48, 116, 52, 54, or 56, include use as markers for tissues in which the corresponding protein is preferentially expressed (to identify constitutively expressed proteins or proteins produced at a particular stage of tissue differentiation or stage of development of a disease state); as molecular weight markers on southern gels; as chromosome markers or tags (when labeled, for example with biotin, a radioactive label or a fluorescent label) to identify chromosomes or to map related gene positions; to compare with endogenous DNA sequences in a mammal to identify potential genetic disorders; as probes to hybridize and thus identify, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; as a probe to "subtract-out" known sequences in the

process of discovering other novel nucleic acid molecules; for selecting and making oligomers for attachment to a "gene chip" or other support, to be used, for example, for examination of expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or to elicit  
5 another immune response.

In certain embodiments, a contiguous portion can be about 15, 25, 30, 40, 50, 75, 100, 200, 300, 400, 500, 750, 1000, 1100, 1250, 1500 or more nucleotides in length. In a particular embodiment, the contiguous portion encompasses the signature sequence of a FATP and is about 1080 nucleotides in length.

10 Further methods to obtain nucleic acids encoding FATPs of the FATP family include PCR and variations thereof (e.g., "RACE" PCR and semi-specific PCR methods). Portions of the nucleic acids having a nucleotide sequence set forth in Figures 44A-44C (SEQ ID NO:46), Figures 46A-46B (SEQ ID NO:48), Figures 111A-B (SEQ ID NO:116), Figures 50A-50C (SEQ ID NO:52), Figure 52 (SEQ ID  
15 NO:54) or Figures 54A-54C (SEQ ID NO:56), (especially "flanking sequences" on either side of a coding region) can be used as primers in methods using the polymerase chain reaction, to produce DNA from an appropriate template nucleic acid.

Once a fragment of the FATP gene is generated by PCR, it can be sequenced,  
20 and the sequence of the product can be compared to other DNA sequences, for example, by using the BLAST Network Service at the National Center for Biotechnology Information. The boundaries of the open reading frame can then be identified using semi-specific PCR or other suitable methods such as library screening. Once the 5' initiator methionine codon and the 3' stop codon have been  
25 identified, a PCR product encoding the full-length gene can be generated using genomic DNA as a template, with primers complementary to the extreme 5' and 3' ends of the gene or to their flanking sequences. The full-length genes can then be cloned into expression vectors for the production of functional proteins.

The invention also relates to isolated proteins or polypeptides such as those  
30 encoded by nucleic acids of the present invention. Isolated proteins can be purified from a natural source or can be made recombinantly. Proteins or polypeptides

referred to herein as "isolated" are proteins or polypeptides that exist in a state different from the state in which they exist in cells in which they are normally expressed in an organism, and include proteins or polypeptides obtained by methods described herein, similar methods or other suitable methods, and also include

5 essentially pure proteins or polypeptides, proteins or polypeptides produced by chemical synthesis or by combinations of biological and chemical methods, and recombinant proteins or polypeptides which are isolated. Thus, the term "isolated" as used herein, indicates that the polypeptide in question exists in a physical milieu distinct from that in which it occurs in nature. Thus, "isolated" includes existing in

10 membrane fragments and vesicles membrane fractions, liposomes, lipid bilayers and other artificial membrane systems. An isolated FATP may be substantially isolated with respect to the complex cellular milieu in which it naturally occurs, and may even be purified essentially to homogeneity, for example as determined by PAGE or column chromatography (for example, HPLC), but may also have further cofactors or

15 molecular stabilizers, such as detergents, added to the purified protein to enhance activity. In one embodiment, proteins or polypeptides are isolated to a state at least about 75% pure; more preferably at least about 85% pure, and still more preferably at least about 95% pure, as determined by Coomassie blue staining of proteins on SDS-polyacrylamide gels. Proteins or polypeptides referred to herein as "recombinant" are

20 proteins or polypeptides produced by the expression of recombinant nucleic acids.

In a preferred embodiment, an isolated polypeptide comprising a FATP, a functional portion thereof, or a functional equivalent of the FATP, has at least one function characteristic of a FATP, for example, transport activity, binding function (e.g., a domain which binds to AMP), or antigenic function (e.g., binding of

25 antibodies that also bind to a naturally-occurring FATP, as that function is found in an antigenic determinant). Functional equivalents can have activities that are quantitatively similar to, greater than, or less than, the reference protein. These proteins include, for example, naturally occurring FATPs that can be purified from tissues in which they are produced (including polymorphic or allelic variants),

30 variants (e.g., mutants) of those proteins and/or portions thereof. Such variants include mutants differing by the addition, deletion or substitution of one or more



amino acid residues, or modified polypeptides in which one or more residues are modified, and mutants comprising one or more modified residues. Portions or fragments of a FATP can range in size from four amino acid residues to the entire amino acid sequence minus one amino acid and include contiguous portions or  
5 fragments about 4, 5, 6, 7, 8, 9, 10, 15, 25, 30, 40, 50, 75, 100, 150, 200, 300, 400, 500, 600 or more amino acid residues in length. In one particular embodiment, the portion or fragment includes the signature sequence of a FATP polypeptide and is about 360 amino acid residues in length.

The isolated proteins of the invention preferably include mammalian fatty  
10 acid transport proteins of the FATP family of homologous proteins. In one embodiment, the extent of amino acid sequence similarity between a polypeptide having one of the amino acid sequences shown in Figure 45 (SEQ ID NO:47), Figure 47 (SEQ ID NO:49), Figure 112 (SEQ ID NO:117), Figure 51 (SEQ ID NO:53), Figure 53 (SEQ ID NO:55), or Figure 55 (SEQ ID NO:57), and the respective  
15 functional equivalents of these polypeptides is at least about 88%. In other embodiments, the degree of amino acid sequence similarity between a FATP and its respective functional equivalent is at least about 91%, at least about 94%, or at least about 97%.

The polypeptides of the invention also include those FATPs encoded by  
20 polynucleotides which are orthologous to those polynucleotides, the sequences of which are described herein in whole or in part. FATPs which are orthologs to those described herein by amino acid sequence, in whole or in part, are, for example, fatty acid transport proteins 1-6 of dog, rat, chimpanzee, monkey, rabbit, guinea pig, baboon and pig, and are also embodiments of the invention.

25 To determine the percent identity or similarity of two amino acid sequences or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment, and non-homologous (dissimilar) sequences can be disregarded for comparison purposes). In a preferred  
30 embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more

preferably at least 60%, and even more preferably at least 70%, 80%, or 90% of the length of the reference sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or  
5 nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein, amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "similarity"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which  
10 need to be introduced for optimal alignment of the two sequences.

The invention also encompasses polypeptides having a lower degree of identity but having sufficient similarity so as to perform one or more of the same functions performed by the polypeptides described herein by amino acid sequence. Similarity for a polypeptide is determined by conserved amino acid substitution.  
15 Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Conservative substitutions are likely to be phenotypically silent. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues  
20 Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe and Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent is found in Bowie *et al.*, *Science* 247:1306-1310 (1990).

TABLE 1. Conservative Amino Acid Substitutions

	Aromatic		Phenylalanine	
			Tryptophan	
			Tyrosine	
	Hydrophobic		Leucine	
			Isoleucine	
			Valine	
	Polar		Glutamine	
			Asparagine	
5	Basic		Arginine	
			Lysine	
			Histidine	
	Acidic		Aspartic Acid	
			Glutamic Acid	
	Small		Alanine	
			Serine	
			Threonine	
			Methionine	
			Glycine	

The comparison of sequences and determination of percent identity and similarity between two sequences can be accomplished using a mathematical algorithm. (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part 1*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and

15 *Sequence Analysis Primer*, Gribskov, M. and Devereaux, J., eds., M. Stockton

Press, New York, 1991). In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at <http://www.gcg.com>), using  
5 either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (Devereux, J., *et al.*, *Nucleic Acids Res.* 12(1):387 (1984)) (available at <http://www.gcg.com>), using a  
10 NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Meyers and W. Miller (*CABIOS*, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length  
15 penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.* (*J. Mol.*  
20 *Biol.* 215:403-10 (1990)). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, word length = 12 to obtain nucleotide sequences homologous to (with calculatably significant similarity to) the nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, word length = 3 to obtain amino acid sequences homologous to  
25 the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.*, (*Nucleic Acids Res.* 25(17):3389-3402 (1997)). When utilizing BLAST and gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See <http://www.ncbi.nlm.nih.gov>. Similarity for nucleotide and  
30 amino acid sequences can be defined in terms of the parameters set by the Advanced Blast search available from NCBI (the National Center for Biotechnology

Information; see, for Advanced BLAST page, [www.ncbi.nlm.nih.gov/cgi-bin/BLAST/nph-newblast?Jform=1](http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST/nph-newblast?Jform=1)). These default parameters, recommended for a query molecule of length greater than 85 amino acid residues or nucleotides have been set as follows: gap existence cost, 11, per residue gap cost, 1; lambda ratio, 0.85. Further explanation of version 2.0 of BLAST can be found on related website pages and in Altschul, S.F. *et al.*, *Nucleic Acids Res.* 25:3389-3402 (1997).

In certain embodiments, a contiguous portion can be about 4, 5, 6, 7, 8, 9, 10, 15, 25, 30, 40, 50, 75, 100, 150, 200, 300, 400, 500, 600 or more amino acid residues in length. In one particular embodiment, the portion or fragment includes the signature sequence of a FATP polypeptide and is about 360 amino acid residues in length.

The invention further relates to fusion proteins, comprising a FATP or functional portion thereof (as described above) as a first moiety, linked to a second moiety not occurring in the FATP as found in nature. Thus, the second moiety can be an amino acid, peptide or polypeptide. The first moiety can be in an N-terminal location, C-terminal location or internal to the fusion protein. In one embodiment, the fusion protein comprises a FATP as the first moiety, and a second moiety comprising a linker sequence and an affinity ligand. Fusion proteins can be produced by a variety of methods. For example, a fusion protein can be produced by the insertion of a FATP gene or portion thereof into a suitable expression vector, such as Bluescript SK +/- (Stratagene, La Jolla, CA), pGEX-4T-2 (Pharmacia, Peapack, NJ), pET-24(+) (Novagen, Madison, WI), or vectors of similar construction. The resulting construct can be introduced into a suitable host cell for expression. Upon expression, fusion protein can be purified from cells by means of a suitable affinity matrix (See e.g., *Current Protocols in Molecular Biology*, Ausubel, F.M. *et al.*, eds., Vol. 2, pp. 16.4.1-16.7.8, containing supplements up through Supplement 42, 1998).

The invention also relates to enzymatically produced, synthetically produced, or recombinantly produced portions of a fatty acid transport protein. Portions of a FATP can be made which have full or partial function on their own, or which when mixed together (though fully, partially, or nonfunctional alone), spontaneously

assemble with one or more other polypeptides to reconstitute a functional protein having at least one function characteristic of a FATP.

Fragments of a FATP can be produced by direct peptide synthesis, for example those using solid-phase techniques (Roberge, J.Y. *et al.*, *Science* 269:202-204 (1995); Merrifield, J., *J. Am. Chem. Soc.* 85:2149-2154 (1963)). Protein synthesis can be performed using manual techniques or by automation. Automated synthesis can be carried out using, for instance, an Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Various fragments of a FATP can be synthesized separately and combined using chemical methods.

One aspect of the invention is a peptide or polypeptide having the amino acid sequence of a portion of a fatty acid transport protein which is hydrophilic rather than hydrophobic, and ordinarily can be detected as facing the outside of the cell membrane. Such a peptide or polypeptide can be thought of as being an extracellular domain of the FATP, or a mimetic of said extracellular domain. It is known, for example, that a portion of human FATP4 that includes a highly conserved motif is involved in AMP-CoA binding function (Stuhlsatz-Krouper, S.M. *et al.*, *J. Biol. Chem.* 44:28642-28650 (1998)).

The term "mimetic" as used herein, refers to a molecule, the structure of which is developed from knowledge of the structure of the FATP of interest, or one or more portions thereof, and, as such, is able to effect some or all of the functions of a FATP.

Portions of a FATP can be prepared by enzymatic cleavage of the isolated protein, or can be made by chemical synthesis methods. Portions of a FATP can also be made by recombinant DNA methods in which restriction fragments, or fragments that may have undergone further enzymatic processing, or synthetically made DNAs are joined together to construct an altered FATP gene. The gene can be made such that it encodes one or more desired portions of a FATP. These portions of FATP can be entirely homologous to a known FATP, or can be altered in amino acid sequence relative to naturally occurring FATPs to enhance or introduce desired properties such as solubility, stability, or affinity to a ligand. A further feature of the gene can be a sequence encoding an N-terminal signal peptide directed to the plasma membrane.

An extracellular domain can be determined by a hydrophobicity plot, such as those shown in Figures 28A, 29A, and 35A, or by a hydrophilicity plot such as those shown in Figures 28C, 29C, 35C, 91, 92 and 93. A polypeptide or peptide comprising all or a portion of a FATP extracellular domain can be used in a pharmaceutical composition. When administered to a mammal by an appropriate route, the polypeptide or peptide can bind to fatty acids and compete with the native FATPs in the membrane of cells, thereby making fewer fatty acid molecules available as substrates for transport into cells, and reducing the amount of fatty acids taken up by, for example, the heart, in the case of FATP6.

Another aspect of the invention relates to a method of producing a fatty acid transport protein, variants or portions thereof, and to expression systems and host cells containing a vector appropriate for expression of a fatty acid transport protein.

Cells that express a FATP, a variant or a portion thereof, or an ortholog of a FATP described herein by amino acid sequence, can be made and maintained in culture, under conditions suitable for expression, to produce protein in the cells for cell-based assays, or to produce protein for isolation. These cells can be procaryotic or eucaryotic. Examples of procaryotic cells that can be used for expression include *Escherichia coli*, *Bacillus subtilis* and other bacteria. Examples of eucaryotic cells that can be used for expression include yeasts such as *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pichia pastoris* and other lower eucaryotic cells, and cells of higher eucaryotes such as those from insects and mammals, such as primary cells and cell lines such as CHO, HeLa, 3T3 and BHK cells, preferably COS cells and human kidney 293 cells, and more preferably Jurkat cells. (See, e.g., Ausubel, F.M. *et al.*, eds. *Current Protocols in Molecular Biology*, Greene Publishing Associates and John Wiley & Sons, Inc., containing Supplements up through Supplement 42, 1998)).

In one embodiment, host cells that produce a recombinant FATP, or a portion thereof, a variant, or an ortholog of a FATP described herein by amino acid sequence, can be made as follows. A gene encoding a FATP, variant or a portion thereof can be inserted into a nucleic acid vector, e.g., a DNA vector, such as a plasmid, phage, cosmid, phagemid, virus, virus-derived vector (e.g., SV40, vaccinia, adenovirus, fowl

pox virus, pseudorabies viruses, retroviruses) or other suitable replicon, which can be present in a single copy or multiple copies, or the gene can be integrated in a host cell chromosome. A suitable replicon or integrated gene can contain all or part of the coding sequence for a FATP or variant, operably linked to one or more expression control regions whereby the coding sequence is under the control of transcription signals and linked to appropriate translation signals to permit translation. The vector can be introduced into cells by a method appropriate to the type of host cells (e.g., transfection, electroporation, infection). For expression from the FATP gene, the host cells can be maintained under appropriate conditions (e.g., in the presence of inducer, normal growth conditions, etc.). Proteins or polypeptides thus produced can be recovered (e.g., from the cells, as in a membrane fraction, from the periplasmic space of bacteria, from culture medium) using suitable techniques. Appropriate membrane targeting signals may be incorporated into the expressed polypeptide. These signals may be endogenous to the polypeptide or they may be heterologous signals.

Polypeptides of the invention can be recovered and purified from cell cultures (or from their primary cell source) by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and high performance liquid chromatography. Known methods for refolding protein can be used to regenerate active conformation if the polypeptide is denatured during isolation or purification.

In a further aspect of the invention are methods for assessing the transport function of any of the fatty acid transport proteins or polypeptides described herein, including orthologs, and in variations of these, methods for identifying an inhibitor (or an enhancer) of such function and methods for assessing the transport function in the presence of a candidate inhibitor or a known inhibitor.

A variety of systems comprising living cells can be used for these methods. Cells to be used in fatty acid transport assays, and further in methods for identifying an inhibitor or enhancer of this function, express one or more FATPs. See Examples



3, 6, 9, 12 and 14 for data on tissue distribution of expression of FATPs, and Examples 10 and 11 describing recombinant cells expressing FATP. Cells for use in cell-based assays described herein can be drawn from a variety of sources, such as isolated primary cells of various organs and tissues wherein one or more FATPs are naturally expressed. In some cases, the cells can be from adult organs, and in some cases, from embryonic or fetal organs, such as heart, lung, liver, intestine, skeletal muscle, kidney and the like. Cells for this purpose can also include cells cultured as fragments of organs or in conditions simulating the cell type and/or tissue organization of organs, in which artificial materials may be used as substrates for cell growth. Other types of cells suitable for this purpose include cells of a cell strain or cell line (ordinarily comprising cells considered to be "transformed") transfected to express one or more FATPs.

A further embodiment of the invention is a method for detecting, in a sample of cells, a fatty acid transport protein, a portion or fragment thereof, a fusion protein comprising a FATP or a portion thereof, or an ortholog as described herein, wherein the cells can be, for instance, cells of a tissue, primary culture cells, or cells of a cell line, including cells into which nucleic acid has been introduced. The method comprises adding to the sample an agent that specifically binds to the protein, and detecting the agent specifically bound to the protein. Appropriate washing steps can be added to reduce nonspecific binding to the agent. The agent can be, for example, an antibody, a ligand or a substrate mimic. The agent can have incorporated into it, or have bound to it, covalently or by high affinity non-covalent interactions, for instance, a label that facilitates detection of the agent to which it is bound, wherein the label can be, but is not limited to, a phosphorescent label, a fluorescent label, a biotin or avidin label, or a radioactive label. The means of detection of a fatty acid transport protein can vary, as appropriate to the agent and label used. For example, for an antibody that binds to the fatty acid transport protein, the means of detection may call for binding a second antibody, which has been conjugated to an enzyme, to the antibody which binds the fatty acid transport protein, and detecting the presence of the second antibody by means of the enzymatic activity of the conjugated enzyme.

Similar principles can also be applied to a cell lysate or a more purified preparation of proteins from cells that may comprise a fatty acid transport protein of interest, for example in the methods of immunoprecipitation, immunoblotting, immunoaffinity methods, that in addition to detection of the particular FATP, can also be used in purification steps, and qualitative and quantitative immunoassays. See, for instance, chapters 11 through 14 in *Antibodies: A Laboratory Manual*, E. Harlow and D. Lane, eds., Cold Spring Harbor Laboratory, 1988.

Isolated fatty acid transport protein or, an antigenically similar portion thereof, especially a portion that is soluble, can be used in a method to select and identify molecules which bind specifically to the FATP. Fusion proteins comprising all of, or a portion of, the fatty acid transport protein linked to a second moiety not occurring in the FATP as found in nature, can be prepared for use in another embodiment of the method. Suitable fusion proteins for this purpose include those in which the second moiety comprises an affinity ligand (e.g., an enzyme, antigen, epitope). FATP fusion proteins can be produced by the insertion of a gene encoding the FATP or a variant thereof, or a suitable portion of such gene into a suitable expression vector, which encodes an affinity ligand (e.g., pGEX-4T-2 and pET-15b, encoding glutathione S-transferase and His-Tag affinity ligands, respectively). The expression vector can be introduced into a suitable host cell for expression. Host cells are lysed and the lysate, containing fusion protein, can be bound to a suitable affinity matrix by contacting the lysate with an affinity matrix. In a particular embodiment, a nucleic acid encodes a portion of a FATP polypeptide which includes a motif or domain, for example, a lipocalin domain or an AMP-binding domain. Such a polypeptide portion can be a functional portion of a FATP protein. The term "lipocalin domain" is an art recognized term and as used herein refers to a particular domain present in FATP proteins. This domain is described as including regions of sequence homology as well as a common tertiary structure represented as an eight stranded antiparallel beta-barrel. (see Banaszak, L. *et al.*, *Advances in Protein Chemistry*, 45: 89-151). Many lipocalin domains can be identified structurally as a sequence contained within the general formula: [DENG]-X-[DENQGSTARK]-X(0,2)-[DENQARK]-[LIVFY]-{CP}-G-{C}-W-[FYWLRH-X]-[LIVMTA], e.g., the lipocalin signature sequence or

consensus pattern (SEQ ID NO: 125). One skilled in the art will recognize that a lipocalin domain for a particular FATP protein can vary in sequence from this general formula. A FATP lipocalin domain can be, for example, identical to the lipocalin signature sequence, or, can exhibit 60, 65, 70, 75, 80, 85, 90, 95 or greater sequence  
5 percent identity in comparison to the general formula provided that it still retains the necessary lipocalin binding function.

For example, a lipocalin domain for each of the human FATPs, hsFATP1 (SEQ ID NO: 126), hsFATP2 (SEQ ID NO: 127), hsFATP3 (SEQ ID NO: 128), hsFATP4 (SEQ ID NO: 129), hsFATP5 (SEQ ID NO: 130), and hsFATP6 (SEQ ID  
10 NO: 131) has been identified. These particular lipocalin domains are located near the N-terminal portion of the specified proteins (see Figure 118). The sequences of these lipocalin domains are highly conserved across the FATP family. A search using the lipocalin signature sequence conducted on a public database ([www.ebi.ac.uk/interpro/](http://www.ebi.ac.uk/interpro/)), indicated that the lipocalin domains of hsFATP1 and  
15 hsFATP4 share identity with signature sequence. In addition, a search directed to identifying sequences having at least 80% identity to the lipocalin signature sequence identified three additional human FATPs, hsFATP3, hsFATP5 and hsFATP6.

A lipocalin domain can also be identified functionally since, for example, it has been identified as a binding motif capable of binding fatty acids. In particular,  
20 the studies described in Experiment 20 demonstrated that fusion proteins including the lipocalin domains from hsFATP4 bound long chain fatty acids such as oleates and palmitates with great specificity. Other fatty acids can also be used to assess binding in FATP4 and other members of the FATP family.

Polypeptides, including fusion polypeptides, which contain a lipocalin domain  
25 can also include additional components. For example, fusion polypeptides containing a lipocalin domain can include amino acid residues from the portion of the protein which is located upstream, *i. e.*, in the direction of the N-terminal end of a FATP protein, from the lipocalin domain. As the term "upstream sequences" is used herein in relation to the lipocalin domain, it is intended to refer to the amino acid residues of  
30 a FATP protein which are located between the signal peptide (when one is present) and the lipocalin domain. In the absence of a signal peptide, the term refers to the

portion of a FATP protein between the lipocalin domain and the amino terminus (see Figure 115).

Fusion polypeptides which contain a lipocalin domain can also include additional domains or motifs, for example, an AMP binding domain can be included.

- 5 For example, an AMP binding domain for each of the human FATPs, hsFATP1 (SEQ ID NO: 132), hsFATP2 (SEQ ID NO: 133), hsFATP3 (SEQ ID NO: 134), hsFATP4 (SEQ ID NO: 135), hsFATP5 (SEQ ID NO: 136) and hsFATP6 (SEQ ID NO: 137) has been identified (see Figure 118).

- 10 In one embodiment, the fusion protein can be immobilized on a suitable affinity matrix under conditions sufficient to bind the affinity ligand portion of the fusion protein to the matrix, and is contacted with one or more candidate binding agents (e.g., a mixture of peptides) to be tested, under conditions suitable for binding of the binding agents to the FATP portion of the bound fusion protein. Next, the affinity matrix with bound fusion protein can be washed with a suitable wash buffer
- 15 to remove unbound candidate binding agents and non-specifically bound candidate binding agents. Those agents which remain bound can be released by contacting the affinity matrix with fusion protein bound thereto with a suitable elution buffer. Wash buffer can be formulated to permit binding of the fusion protein to the affinity matrix, without significantly disrupting binding of specifically bound binding agents. In this
- 20 aspect, elution buffer can be formulated to permit retention of the fusion protein by the affinity matrix, but can be formulated to interfere with binding of the candidate binding agents to the target portion of the fusion protein. For example, a change in the ionic strength or pH of the elution buffer can lead to release of specifically bound agent, or the elution buffer can comprise a release component or components
- 25 designed to disrupt binding of specifically bound agent to the target portion of the fusion protein.

- Immobilization can be performed prior to, simultaneous with, or after, contacting the fusion protein with candidate binding agent, as appropriate. Various permutations of the method are possible, depending upon factors such as the
- 30 candidate molecules tested, the affinity matrix-ligand pair selected, and elution buffer formulation. For example, after the wash step, fusion protein with binding agent

molecules bound thereto can be eluted from the affinity matrix with a suitable elution buffer (a matrix elution buffer, such as glutathione for a GST fusion). Where the fusion protein comprises a cleavable linker, such as a thrombin cleavage site, cleavage from the affinity ligand can release a portion of the fusion with the candidate agent bound thereto. Bound agent molecules can then be released from the fusion protein or its cleavage product by an appropriate method, such as extraction.

One or more candidate binding agents can be tested simultaneously. Where a mixture of candidate binding agents is tested, those found to bind by the foregoing processes can be separated (as appropriate) and identified by suitable methods (e.g., PCR, sequencing, chromatography). Large libraries of candidate binding agents (e.g., peptides, RNA oligonucleotides) produced by combinatorial chemical synthesis or by other methods can be tested (see e.g., Ohlmeyer, M.H.J. *et al.*, *Proc. Natl. Acad. Sci. USA* 90:10922-10926 (1993) and DeWitt, S.H. *et al.*, *Proc. Natl. Acad. Sci. USA* 90:6909-6913 (1993), relating to tagged compounds; see also Rutter, W.J. *et al.* U.S. Patent No. 5,010,175; Huebner, V.D. *et al.*, U.S. Patent No. 5,182,366; and Geysen, H.M., U.S. Patent No. 4,833,092). Random sequence RNA libraries (see Ellington, A.D. *et al.*, *Nature* 346:818-822 (1990); Bock, L.C. *et al.*, *Nature* 355:584-566 (1992); and Szostak, J.W., *Trends in Biochem. Sci.* 17:89-93 (March, 1992)) can also be screened according to the present method to select RNA molecules which bind to a target FATP or FATP fusion protein. Where binding agents selected from a combinatorial library by the present method carry unique tags, identification of individual biomolecules by chromatographic methods is possible. Where binding agents do not carry tags, chromatographic separation, followed by mass spectrometry to ascertain structure, can be used to identify binding agents selected by the method, for example.

The invention also comprises a method for identifying an agent which inhibits interaction between a fatty acid transport protein (e.g., one comprising the amino acid sequence in SEQ ID NO:47, SEQ ID NO:49, SEQ ID NO:117, SEQ ID NO:53, SEQ ID NO:55, or SEQ ID NO:57), and a ligand of said protein. The FATP can be one described by an amino acid sequence herein, a portion or fragment thereof, a variant thereof, or an ortholog thereof, or a FATP fusion protein. Here, a ligand can be, for

instance, a substrate, or a substrate mimic, an antibody, or a compound, such as a peptide, that binds with specificity to a site on the protein. The method comprises combining, not limited to a particular order, the fatty acid protein, the ligand of the protein, and a candidate agent to be assessed for its ability to inhibit interaction  
5 between the protein and the ligand, under conditions appropriate for interaction between the protein and the ligand (e.g., pH, salt, temperature conditions conducive to appropriate conformation and molecular interactions); determining the extent to which the protein and ligand interact; and comparing (1) the extent of protein-ligand interaction in the presence of candidate agent with (2) the extent of protein-ligand  
10 interaction in the absence of candidate agent, wherein if (1) is less than (2), then the candidate agent is one which inhibits interaction between the protein and the ligand.

The method can be facilitated, for example, by using an experimental system which employs a solid support (column chromatography matrix, wall of a plate, microtiter wells, column pore glass, pins to be submerged in a solution, beads, etc.)  
15 to which the protein can be attached. Accordingly, in one embodiment, the protein can be fixed to a solid phase directly or indirectly, by a linker. The candidate agent to be tested is added under conditions conducive for interaction and binding to the protein. The ligand is added to the solid phase system under conditions appropriate for binding. Excess ligand is removed, as by a series of washes done under  
20 conditions that do not disrupt protein-ligand interactions. Detection of bound ligand can be facilitated by using a ligand that carries a label (e.g., fluorescent, chemiluminescent, radioactive). In a control experiment, protein and ligand are allowed to interact in the absence of any candidate agent, under conditions otherwise identical to those used for the "test" conditions where candidate inhibiting agent is  
25 present, and any washes used in the test conditions are also used in the control. The extent to which ligand binds to the protein in the presence of candidate agent is compared to the extent to which ligand binds to the protein in the absence of the candidate agent. If the extent to which interaction of the protein and the ligand occurs is less in the presence of the candidate agent than in the absence of the  
30 candidate agent, the candidate agent is an agent which inhibits interaction between the protein and the ligand of the protein.

In a further embodiment, an inhibitor (or an enhancer) of a fatty acid transport protein can be identified. The method comprises steps which are, or are variations of the following: contacting the cells with fatty acid, wherein the fatty acid can be labeled for convenience of detection; contacting a first aliquot of the cells with an agent being tested as an inhibitor (or enhancer) of fatty acid uptake while maintaining a second aliquot of cells under the same conditions but without contact with the agent; and measuring (e.g., quantitating) fatty acid in the first and second aliquots of cells; wherein a lesser quantity of fatty acid in the first aliquot compared to that in the second aliquot is indicative that the agent is an inhibitor of fatty acid uptake by a fatty acid transport protein. A greater quantity of fatty acid in the first aliquot compared to that in the second aliquot is indicative that the agent is an enhancer of fatty acid uptake by a fatty acid transport protein.

A particular embodiment of identifying an inhibitor or enhancer of fatty acid transport function employs the above steps, but also employs additional steps preceding those given above: introducing into cells of a cell strain or cell line ("host cells" for the intended introduction of, or after the introduction of, a vector) a vector comprising a fatty acid transport protein gene, wherein expression of the gene can be regulatable or constitutive, and providing conditions to the host cells under which expression of the gene can occur.

The terms "contacting" and "combining" as used herein in the context of bringing molecules into close proximity to each other, can be accomplished by conventional means. For example, when referring to molecules that are soluble, contacting is achieved by adding the molecules together in a solution. "Contacting" can also be adding an agent to a test system, such as a vessel containing cells in tissue culture.

The term "inhibitor" or "antagonist", as used herein, refers to an agent which blocks, diminishes, inhibits, hinders, limits, decreases, reduces, restricts or interferes with fatty acid transport into the cytoplasm of a cell, or alternatively and additionally, prevents or impedes the cellular effects associated with fatty acid transport. The term "enhancer" or "agonist", as used herein, refers to an agent which augments, enhances, or increases fatty acid transport into the cytoplasm of a cell. An antagonist will

decrease fatty acid concentration, fatty acid metabolism and byproduct levels in the cell, leading to phenotypic and molecular changes.

In order to produce a "host cell" type suitable for fatty acid uptake assays and for assays derived therefrom for identifying inhibitors or enhancers thereof, a nucleic acid vector can be constructed to comprise a gene encoding a fatty acid transport protein, for example, human FATP1, FATP2, FATP3, FATP4, FATP5, FATP6, a mutant or variant thereof, an ortholog of the human proteins, such as mouse orthologs or orthologs found in other mammals, or a FATP family protein of origin in an organism other than a mammal. The gene of the vector can be regulatable, such as by the placement of the gene under the control of an inducible or repressible promoter in the vector (e.g., inducible or repressible by a change in growth conditions of the host cell harboring the vector, such as addition of inducer, binding or functional removal of repressor from the cell milieu, or change in temperature) such that expression of the FATP gene can be turned on or initiated by causing a change in growth conditions, thereby causing the protein encoded by the gene to be produced, in host cells comprising the vector, as a plasma membrane protein. Alternatively, the FATP gene can be constitutively expressed.

A vector comprising a FATP gene, such as a vector described herein, can be introduced into host cells by a means appropriate to the vector and to the host cell type. For example, commonly used methods such as electroporation, transfection, for instance, transfection using  $\text{CaCl}_2$ , and transduction (as for a virus or bacteriophage) can be used. Host cells can be, for example, mammalian cells such as primary culture cells or cells of cell lines such as COS cells, 293 cells or Jurkat cells. Host cells can also be, in some cases, cells derived from insects, cells of insect cell lines, bacterial cells, such as *E. coli*, or yeast cells, such as *S. cerevisiae*. It is preferred that the fatty acid transport protein whose function is to be assessed, with or without a candidate inhibitor or enhancer, be produced in host cells whose ancestor cells originated in a species related to the species of origin of the FATP gene encoding the fatty acid transport protein. For example, it is preferable that tests of function or of inhibition or enhancement of a mammalian FATP be carried out in host mammalian cells producing the FATP, rather than bacterial cells or yeast cells.



Host cells comprising a vector comprising a regulatable FATP gene can be treated so as to allow expression of the FATP gene and production of the encoded protein (e.g., by contacting the cells with an inducer compound that effects transcription from an inducible promoter operably linked to the FATP gene).

5           Alternatively, host cells containing an endogenous FATP gene can be engineered to activate or deactivate expression of the FATP gene and production of the encoded protein. For example, homologous recombination, often referred to as targeting, can be utilized to alter the regulatory region associated with the FATP gene to increase or decrease the level of expression. Alteration of the regulatory region  
10       can include disablement of the regulatory region associated with the FATP gene and/or replacement of the region or a portion of the region. A variety of regulatory regions are known which can be transfected into cells to cause an endogenous gene to display a pattern of induction or expression that differs from that of the cell prior to transfection.

15           The test agent (e.g., an agonist or antagonist) is added to the cells to be used in a fatty acid transport assay, in the presence or absence of test agent, under conditions suitable for production and/or maintenance of the expressed FATP in a conformation appropriate for association of the FATP with test agent and substrate. For example, conditions under which an agent is assessed, such as media and  
20       temperature requirements, can, initially, be similar to those necessary for transport of typical fatty acid substrates across the plasma membrane. One of ordinary skill in the art will know how to vary experimental conditions depending upon the biochemical nature of the test agent. The test agent can be added to the cells in the presence of fatty acid, or in the absence of fatty acid substrate, with the fatty acid substrate being  
25       added following the addition of the test agent. The concentration at which the test agent can be evaluated can be varied, as appropriate, to test for an increased effect with increasing concentrations.

          Test agents to be assessed for their effects on fatty acid transport can be any chemical (element, molecule, compound), made synthetically, made by recombinant  
30       techniques or isolated from a natural source. For example, test agents can be peptides, polypeptides, peptoids, sugars, hormones, or nucleic acid molecules, such as

antisense nucleic acid molecules. In addition, test agents can be small molecules or molecules of greater complexity made by combinatorial chemistry, for example, and compiled into libraries. These libraries can comprise, for example, alcohols, alkyl halides, amines, amides, esters, aldehydes, ethers and other classes of organic compounds. Test agents can also be natural or genetically engineered products isolated from lysates of cells, bacterial, animal or plant, or can be the cell lysates themselves. Presentation of test compounds to the test system can be in either an isolated form or as mixtures of compounds, especially in initial screening steps.

Thus, the invention relates to a method for identifying agents which alter fatty acid transport, the method comprising providing the test agent to the cell (wherein "cell" includes the plural, and can include cells of a cell strain, cell line or culture of primary cells or organ culture, for example), under conditions suitable for binding to its target, whether to the FATP itself or to another target on or in the cell, wherein the transformed cell comprises a FATP.

In greater detail, to test one or more agents or compounds (e.g., a mixture of compounds can conveniently be screened initially) for inhibition of the transport function of a fatty acid transport protein, the agent(s) can be contacted with the cells. The cells can be contacted with a labeled fatty acid. The fatty acid can be, for example, a known substrate of the fatty acid transport protein such as oleate or palmitate. The fatty acid can itself be labeled with a radioactive isotope, (e.g.,  $^3\text{H}$  or  $^{14}\text{C}$ ) or can have a radioactively labeled adduct attached. In other variations, the fatty acid can have chemically attached to it a fluorescent label, or a substrate for an enzyme occurring within the cells, wherein the substrate yields a detectable product, such as a highly colored or fluorescent product. Addition of candidate inhibitors and labeled substrate to the cells comprising fatty acid transport protein can be in either order or can be simultaneous.

A second aliquot of cells, which can be called "control" cells (a "first" aliquot of cells can be called "test" cells), is treated, if necessary (as in the case of transformed "host" cells), so as to allow expression of the FATP gene, and is contacted with the labeled substrate of the fatty acid transport protein. The second aliquot of cells is not contacted with one or more agents to be tested for inhibition of

the transport function of the protein produced in the cells, but is otherwise kept under the same culture conditions as the first aliquot of cells.

In a further step of a method to identify inhibitors of a fatty acid transport protein, the labeled fatty acid is measured in the first and second aliquots of cells. A preliminary step of this measurement process can be to separate the external medium from the cells so as to be able to distinguish the labeled fatty acid external to the cells from that which has been transported inside the cells. This can be accomplished, for instance, by removing the cells from their growth container, centrifuging the cell suspension, removing the supernatant and performing one or more wash steps to extensively dilute the remaining medium which may contain labeled fatty acid. Detection of the labeled fatty acid can be by a means appropriate to the label used. For example, for a radioactive label, detection can be by scintillation counting of appropriately prepared samples of cells (e.g., lysates or protein extracts); for a fluorescent label, by measuring fluorescence in the cells by appropriate instrumentation.

If a compound tested as a candidate inhibitor of transport function causes the test cells to have less labeled fatty acid detected in the cells than that detected in the control cells, then the compound is an inhibitor of the fatty acid transport protein. Procedures analogous to those above can be devised for identifying enhancers (agonists of FATPs) of fatty acid transport function wherein if the test cells contain more labeled fatty acid than that detected in the control cells, or if the fatty acid is taken up at a higher rate, then the compound being tested can be concluded to be an enhancer of the fatty acid transport protein.

Example 13 describes use of an assay of this type to identify an inhibitor of a FATP. In Example 13, an antisense oligonucleotide which specifically inhibits biosynthesis of mmFATP4 was demonstrated to inhibit fatty acid uptake into mouse enterocytes. Similarly, antisense oligonucleotides directed towards specifically inhibiting the biosynthesis of FATP6 in heart cells, FATP5 in liver cells, FATP3 in lung cells, and FATP2 in colon cells, can be demonstrated as examples of "test agents" that inhibit fatty acid transport.

Another assay to determine whether an agent is an inhibitor (or enhancer) of fatty acid transport employs animals, one or more of which are administered the agent, and one or more of which are maintained under similar conditions, but are not administered the agent. Both groups of animals are given fatty acids (e.g., orally, intravenously, by tube inserted into stomach or intestine), and the fatty acids taken up into a bodily fluid (e.g., serum) or into an organ or tissue of interest are measured from comparable samples taken from each group of animals. The fatty acids may carry a label (e.g., radioactive) to facilitate detection and quantitation of fatty acids taken up into the fluid or tissue being sampled. This type of assay can be used alone or can be used in addition to *in vitro* assays of a candidate inhibitor or enhancer.

An agent determined to be an inhibitor (or enhancer) of FATP function, such as fatty acid binding and/or fatty acid uptake, can be administered to cells in culture, or *in vivo*, to a mammal (e.g. human) to inhibit (or enhance) FATP function. Such an agent may be one that acts directly on the FATP (for example, by binding) or can act on an intermediate in a biosynthetic pathway to produce FATP, such as transcription of the FATP gene, processing of the mRNA, or translation of the mRNA. An example of such an agent is antisense oligonucleotide.

Antisense methods similar to those illustrated in Example 13 can be used to determine the target FATP of a compound or agent that has an inhibitory or enhancing effect on fatty acid uptake. For example, antisense oligonucleotide directed to the inhibition of FATP4 biosynthesis can be added to lung cells or cell lines derived from lung cells. In addition, antisense oligonucleotides directed to the inhibition of other FATPs, except for FATP3, can also be added to the lung cells. The administration of antisense oligonucleotides in this manner ensures that the predominant FATP activity remaining in the cells comes from FATP3. After a period of incubation of the cells with the antisense oligonucleotides sufficient to deplete the plasma membrane of the FATPs whose biosynthesis has been inhibited, a test agent, preferably one that has been shown by some preliminary test to have an inhibitory or enhancing activity on fatty acid transport, can be added to the lung cells. If the test agent is now demonstrated, after treatment of the cells with antisense oligonucleotides, to have an inhibitory or enhancing activity on fatty acid transport in

the lung cells, it can be concluded that the target of the test agent is FATP3, or a molecule involved in the biosynthesis or activity of FATP3.

In another type of cell-based assay for uptake of fatty acids, a change of intracellular pH resulting from the uptake of fatty acids can be followed by an  
5 indicator fluorophore. The fluorophore can be taken up by the cells in a preincubation step. Fatty acids can be added to the cell medium, and after some period of incubation to allow FATP-mediated uptake of fatty acids, the change in  $\lambda_{\max}$  of fluorescence can be measured, as an indicator of a change in intracellular pH, as the  $\lambda_{\max}$  of fluorescence of the fluorophore changes with the pH of its environment,  
10 thereby indicating uptake of fatty acids. One such fluorophore is BCECF (2', 7'-bis(2-carboxyethyl)-5(6)-carboxyfluorescein; Rink, T.J. *et al.*, *J. Cell. Biol.* 95: 189 (1982)).

In assays similar to those described above, a candidate inhibitor or enhancer of fatty acid transport function can be added (or mock-added, for control cultures) to  
15 cultures of cells engineered to express a desired FATP to which fatty acid substrate is also added. Inhibition of fatty acid uptake is indicated by a lack of the drop in pH, indicating fatty acid uptake, that is seen in control cells. Enhancement of fatty acid uptake is indicated by a decrease in intracellular pH, as compared to control cells not receiving the candidate enhancer of fatty acid transport function.

20 Yeast cells can be used in a similar cell-based assay for the uptake of fatty acids mediated by a FATP, and such an assay can be adapted to a screening assay for the identification of agents that inhibit or enhance fatty acid uptake by an FATP. Yeast cells lacking an endogenous FATP activity (mutated, disrupted or deleted for *FAT1*; Faergeman, N.J. *et al.*, *J. Biol. Chem.* 272(13):8531-8538 (1997); Watkins, P.A. *et al.*, *J. Biol. Chem.* 273(29):18210-18219 (1998)) can be engineered to harbor  
25 a related gene of the family of FATP-encoding genes, such as a mammalian FATP (e.g., human FATP4).

Examples of expression vectors include pEG (Mitchell, D.A., *et al.*, *Yeast* 9:715-723 (1993)) and pDAD1 and pDAD2, which contain a *GAL1* promoter (Davis,  
30 L. I. and Fink, G. R., *Cell* 61:965-978 (1990)). A variety of promoters are suitable for expression. Available yeast vectors offer a choice of promoters. In one

embodiment, the inducible *GAL1* promoter is used. In another embodiment, the constitutive *ADHI* promoter (alcohol dehydrogenase; Bennetzen, J. L. and Hall, B. D., *J. Biol. Chem.* 257:3026-3031 (1982)) can be used to express an inserted gene on glucose-containing media. An example of a vector suitable for expression of a  
5 heterologous FATP gene in yeast is pQB169.

With the introduced FATP gene providing the only fatty acid transport protein function for the yeast cells, it is possible to study effect of the heterologous FATP on fatty acid transport into the yeast cells in isolation. Assays for the uptake of fatty acids into the yeast cells can be devised that are similar to those described above  
10 and/or those assays that have been illustrated in the Examples. Tests for candidate inhibitors or enhancers of the heterologous FATP can be done in cultures of yeast cells, wherein the yeast cells are incubated with fatty acid substrate and an agent to be tested as an inhibitor or enhancer of FATP function. FATP uptake after a period of time can be measured by analyzing the contents of the yeast cells for fatty acid  
15 substrate, as compared with control yeast cells incubated with the fatty acid, but not with the test agent. Yeast cells have the additional advantage, over mammalian cells in culture, for example, that yeast cells can be forced to rely upon fatty acids as their only source of carbon, if the growth medium supplied to the yeast cells is formulated to contain no other source of carbon. Thus, the effect of the heterologous FATP on  
20 fatty acid uptake and metabolism in the engineered yeast cells can be amplified. An agent that efficiently blocks transport function of the heterologous FATP could result in death of the yeast cells. Thus, in this case, inhibition of function of the heterologous FATP can result in loss of viability. A simple measure of viability is turbidity of the yeast suspension culture, which can be adapted to a high throughput  
25 screening assay for effects of various agents to be tested, using microtiter plates or similar devices for small-volume cultures of the engineered yeast cells.

Cell-free assays can also be used to measure the transport of fatty acids across a membrane, and therefor also to assess a test treatment or test agent for its effect on the rate or extent of fatty acid transport. An isolated FATP, for example in the  
30 presence of a detergent that preserves the native 3-dimensional structure of the FATP, or partially purified FATP, can be used in an artificial membrane system typically

-56-

used to preserve the native conformation and activity of membrane proteins. Such systems include liposomes, artificial bilayers of phospholipids, isolated plasma membrane such as cell membrane fragments, cell membrane fractions, or cell membrane vesicles, and other systems in which the FATP can be properly oriented within the membrane to have transport activity. Assays for transport activity can be performed using methods analogous to those that can be used in cells engineered to predominantly express one FATP whose function is to be measured. A labeled (e.g., radioactively labeled) fatty acid substrate can be incubated with one side of a bilayer or in a suspension of liposomes constructed to integrate a properly oriented FATP.

10 The accumulation of fatty acids with time can be measured, using appropriate means to detect the label (e.g., scintillation counting of medium on each side of the bilayer, or of the contents of liposomes isolated from the surrounding medium). Assays such as these can be adapted to use for the testing of agents which might interact with the FATP to produce an inhibitory or an enhancing effect on the rate or extent of fatty acid transport. That is, the above-described assay can be done in the presence or

15 absence of the agent to be tested, and the results compared.

For examples of isolation of membrane proteins (ADP/ATP carrier and uncoupling protein), reconstitution into phospholipid vesicles, and assays of transport, see Klingenberg, M. *et al.*, *Methods Enzymol.* 260:369-389 (1995). For an example of a membrane protein (phosphate carrier of *Saccharomyces cerevisiae*) that was purified and solubilized from *E. coli* inclusion bodies, see Schroer, A. *et al.*, *J. Biol. Chem.* 273: 14269-14276 (1998). The Glut1 glucose transporter of rat has been expressed in yeast. A crude membrane fraction of the yeast was prepared and reconstituted with soybean phospholipids into liposomes. Glucose transport activity

20 could be measured in the liposomes (Kasahara, T. and Kasahara, M., *J. Biol. Chem.* 273: 29113-29117 (1998)). Similar methods can be applied to the proteins and polypeptides of the invention.

Another embodiment of the invention is a method for inhibiting fatty acid uptake in a mammal (e.g., a human), comprising administering to the mammal a

30 therapeutically effective amount of an inhibitor of the transport function of one or more of the fatty acid transport proteins, thereby decreasing fatty acid uptake by cells

comprising the fatty acid protein(s). Where it is desirable to reduce the uptake of fatty acids, for example, in the treatment of chronic obesity or as a part of a program of weight control or hyperlipidemia control in a human, one or more inhibitors of one or more of the fatty acid transport proteins can be administered in an effective dose, and by an effective route, for example, orally, or by an indwelling device that can deliver doses to the small intestine. The inhibitor can be one identified by methods described herein, or can be one that is, for instance, structurally related to an inhibitor identified by methods described herein (e.g., having chemical adducts to better stabilize or solubilize the inhibitor). The invention further relates to compositions comprising inhibitors of fatty acid uptake in a mammal, which may further comprise pharmaceutical carriers suitable for administration to a subject mammal, such as sterile solubilizing or emulsifying agents.

A further embodiment of the present invention is a method of enhancing or increasing fatty acid uptake, such as enhancing or increasing LCFA uptake in the small intestine (e.g., to treat or prevent a malabsorption syndrome or other wasting condition) or in the liver (e.g., by an enhancer of FATP5 transport activity to treat acute liver failure) or in the kidney (e.g., by an enhancer of FATP2 transport activity to treat kidney failure). In this embodiment, a therapeutically effective amount of an enhancer of the transport function of one or more of the fatty acid transport proteins can be administered to a mammalian subject, with the result that fatty acid uptake in the small intestine is enhanced. In this embodiment, one or more enhancers of one or more of fatty acid transport proteins is administered in an effective dose and by a route (e.g., orally or by a device, such as an indwelling catheter or other device) which can deliver doses to the gut. The enhancer of FATP function (e.g., an enhancer of FATP4 function) can be identified by methods described herein or can be one that is structurally similar to an enhancer identified by methods described herein.

Aerobic reperfusion of ischemic myocardium is a common clinical event which can occur during such treatments as cardiac surgery, angioplasty, and thrombolytic therapy after a myocardial infarction. During reperfusion, a rapid recovery of myocardial energy production is essential for the complete recovery of contractile function. Not only the extent of recovery of myocardial energy



metabolism but also the type of energy substrate used by the heart during reperfusion are important determinants of functional recovery. Circulating fatty acid levels increase following acute myocardial infarction or during cardiac surgery, such that during and following ischemia the heart muscle can be exposed to very high

5 concentrations of fatty acids (Lopaschuk, G.D. and W. C. Stanley, *Science and Medicine* (November/December 1997)). High plasma fatty acid concentrations increase the severity of ischemic damage in a number of experimental models of cardiac ischemia and have been linked to depression of mechanical function during aerobic reperfusion of previously ischemic hearts. Further data show that modifying

10 fatty acid utilization can be beneficial for heart function in ischemia and can be a useful approach for the treatment of angina. See, e.g., Desideri and Celegon, *Am. J. Cardiol.* 82(5A):50K-53K; Lopaschuk, *Am. J. Cardiol.* 82(5A):14K-17K. Plasma fatty acid concentrations can be reduced by administering to a human subject or other mammal an effective amount of an inhibitor of a FATP such as FATP2 or FATP4,

15 thereby providing a way of reducing fatty acid utilization by the heart.

In a further embodiment of the invention, a therapeutically effective amount of an inhibitor of hsFATP6 can be administered to a human patient by a suitable route, to reduce the uptake of fatty acids by cardiac muscle. This treatment is desirable in patients who are diagnosed as having, or who are at risk of, abnormal

20 accumulations of fatty acids in the heart or a detrimentally high rate of uptake of fatty acids into the heart, because of ischemic heart disease, or following ischemia or trauma to the heart.

The invention further relates to antibodies that bind to an isolated or recombinant fatty acid transport protein of the FATP family, including portions of

25 antibodies, which can specifically recognize and bind to one or more FATPs. The antibodies and portions thereof of the invention include those which bind to one or more FATPs of mouse or other mammalian species. In a preferred embodiment, the antibodies specifically bind to a naturally occurring FATP of humans. The antibodies can be used in methods to detect or to purify a protein of the present invention or a

30 portion thereof by various methods of immunoaffinity chromatography, to inhibit the

function of a protein in a method of therapy, or to selectively inactivate an active site, or to study other aspects of the structure of these proteins, for example.

The antibodies of the present invention can be polyclonal or monoclonal. The term antibody is intended to encompass both polyclonal and monoclonal antibodies.

5 Antibodies of the present invention can be raised against an appropriate immunogen, including proteins or polypeptides of the present invention, such as an isolated or recombinant FATP1, FATP2, FATP3, FATP4, FATP5, FATP6, mtFATP, ceFATPa, ceFATPb, scFATP or portions thereof, or synthetic molecules, such as synthetic peptides (e.g., conjugated to a suitable carrier). Preferred embodiments are antibodies  
10 that bind to any of the following: hsFATP1, hsFATP2, hsFATP3, hsFATP4, hsFATP5 or hsFATP6. The immunogen can be a polypeptide comprising a portion of a FATP and having at least one function of a fatty acid transport protein, as described herein.

The term antibody is also intended to encompass single chain antibodies,  
15 chimeric, humanized or primatized (CDR-grafted) antibodies and the like, as well as chimeric or CDR-grafted single chain antibodies, comprising portions from more than one species. For example, the chimeric antibodies can comprise portions of proteins derived from two different species, joined together chemically by conventional techniques or prepared as a single contiguous protein using genetic  
20 engineering techniques (e.g., DNA encoding the protein portions of the chimeric antibody can be expressed to produce a contiguous protein chain. See, e.g., Cabilly et al., U.S. Patent No. 4,816,567; Cabilly *et al.*, European Patent No. 0,125,023 B1; Boss *et al.*, U.S. Patent No. 4,816,397; Boss *et al.*, European Patent No. 0,120,694 B1; Neuberger, M.S. *et al.*, WO 86/01533; Neuberger, M.S. *et al.*, European Patent  
25 No. 0,194,276 B1; Winter, U.S. Patent No. 5,225,539; Winter, European Patent No. 0,239,400 B1; Queen *et al.*, U.S. Patent No. 5,585,089; and Queen *et al.*, European Patent No. EP 0 451 216 B1. See also, Newman, R. *et al.*, *BioTechnology*, 10:1455-1460 (1992), regarding primatized antibody, and Ladner *et al.*, U.S. Patent No. 4,946,778 and Bird, R.E. *et al.*, *Science*, 242:423-426 (1988) regarding single chain  
30 antibodies.)

Whole antibodies and biologically functional fragments thereof are also encompassed by the term antibody. Biologically functional antibody fragments which can be used include those fragments sufficient for binding of the antibody fragment to a FATP to occur, such as Fv, Fab, Fab' and F(ab')<sub>2</sub> fragments. Such  
5 fragments can be produced by enzymatic cleavage or by recombinant techniques. For instance, papain or pepsin cleavage can generate Fab or F(ab')<sub>2</sub> fragments, respectively. Antibodies can also be produced in a variety of truncated forms using antibody genes in which one or more stop codons have been introduced upstream of the natural stop site. For example, a chimeric gene encoding a F(ab')<sub>2</sub> heavy chain  
10 portion can be designed to include DNA sequences encoding the CH<sub>1</sub> domain and hinge region of the heavy chain.

Preparation of immunizing antigen (whole cells comprising FATP on the cell surface or purified FATP), and polyclonal and monoclonal antibody production can be performed using any suitable technique. A variety of methods have been  
15 described (See e.g., Kohler *et al.*, *Nature*, 256: 495-497 (1975) and *Eur. J. Immunol.* 6: 511-519 (1976); Milstein *et al.*, *Nature* 266: 550-552 (1977); Koprowski *et al.*, U.S. Patent No. 4,172,124; Harlow, E. and D. Lane, 1988, *Antibodies: A Laboratory Manual*, (Cold Spring Harbor Laboratory: Cold Spring Harbor, NY); Chapter 11 In *Current Protocols In Molecular Biology*, Vol. 2 (containing supplements up through  
20 Supplement 42, 1998), Ausubel, F.M. *et al.*, eds., (John Wiley & Sons: New York, NY)). Generally, a hybridoma can be produced by fusing a suitable immortal cell line (e.g., a myeloma cell line such as SP2/0) with antibody producing cells. The antibody producing cells, preferably those obtained from the spleen or lymph nodes, can be obtained from animals immunized with the antigen of interest. Immunization  
25 of animals can be by introduction of whole cells comprising fatty acid transport protein on the cell surface. The fused cells (hybridomas) can be isolated using selective culture conditions, and cloned by limiting dilution. Cells which produce antibodies with the desired specificity can be selected by a suitable assay (e.g., ELISA).

30 Other suitable methods of producing or isolating antibodies (including human antibodies) of the requisite specificity can be used, including, for example, methods

which select recombinant antibody from a library (e.g., Hoogenboom *et al.*, WO 93/06213; Hoogenboom *et al.*, U.S. Patent No. 5,565,332; WO 94/13804, published June 23, 1994; and Dower, W.J. *et al.*, U.S. Patent No. 5,427,908), or which rely upon immunization of transgenic animals (e.g., mice) capable of  
5 producing a full repertoire of human antibodies (see e.g., Jakobovits *et al.*, *Proc. Natl. Acad. Sci. USA*, 90: 2551-2555 (1993); Jakobovits *et al.*, *Nature*, 362:255-258 (1993); Lonberg *et al.*, U.S. Patent No. 5,569,825; Lonberg *et al.*, U.S. Patent No. 5,545,806; Surani *et al.*, U.S. Patent No. 5,545,807; and Kucherlapati, R. *et al.*, European Patent No. EP 0 463 151 B1).

10 Another aspect of the invention is a method for directing an agent to cardiac muscle. The differential expression of FATP6 in cardiac muscle but not in other tissue types allows for the specific targeting of drugs, diagnostic agents, tagging labels, histological stains or other substances specifically to cardiac muscle. A targeting vehicle can be used for the delivery of such a substance. Targeting vehicles  
15 which bind specifically to FATP6 can be linked to a substance to be delivered to the cells of cardiac muscle. The linkage can be, for instance, via one or more covalent bonds, or by high affinity non-covalent bonds. A targeting vehicle can be an antibody, for instance, or other compound (e.g., a fatty acid or fatty acid analog) which binds to FATP6 with high specificity.

20 Targeting vehicles specific to the heart-specific protein FATP6 have *in vivo* (e.g., therapeutic and diagnostic) applications. For example, an antibody which specifically binds to FATP6 can be conjugated to a drug to be targeted to the heart (e.g., a cardiac glycoside to treat congestive heart failure, or  $\beta$ -adrenergic agents, sodium channel blockers or calcium channel blockers to treat arrhythmias). A  
25 substance (e.g., a radioactive substance) which can be detected (e.g., a label) *in vivo* can also be linked to a targeting vehicle which specifically binds to a heart-specific protein such as FATP6, and the conjugate can be used as a labeling agent to identify cardiac muscle cells.

Targeting vehicles specific to FATP6 find further applications *in vitro*. For  
30 example, an FATP6-specific targeting vehicle, such as an antibody (a polyclonal preparation or monoclonal) which specifically binds to FATP6, can be linked to a

substance which can be used as a stain for a tissue sample (e.g., horseradish peroxidase) to provide a method for the identification of cardiac muscle in a sample, as can be used in embryology studies, for example.

In a similar manner, an agent can be directed to the liver of a mammal, as  
5 FATP5 is expressed in liver but not in other tissue types. A targeting vehicle which specifically binds to FATP5 can be conjugated to a drug for delivery of the drug to the liver, such as a drug to treat hepatitis, Wilson's disease, lipid storage diseases and liver cancer. As with targeting vehicles specific to FATP6, targeting vehicles specific to FATP5 can be used in studying tissue samples *in vitro*.

10 The invention also relates to compositions comprising a modulator of FATP function. The term "modulate" as used herein refers to the ability of a molecule to alter the function of another molecule. Thus, modulate could mean, for example, inhibit, antagonize, agonize, upregulate, downregulate, induce, or suppress. A modulator has the capability of altering function of its target. Such alteration can be  
15 accomplished at any stage of the transcription, translation, expression or function of the protein, so that, for example, modulation of a target gene can be accomplished by modulation of the DNA or RNA encoding the protein, and the protein itself.

Antagonists or agonists (inhibitors or enhancers) of the FATPs of the invention, antibodies that bind a FATP, or mimetics of a FATP can be employed in  
20 combination with a non-sterile or sterile carrier or carriers for use with cells, tissues or organisms, such as a pharmaceutical carrier suitable for administration to a mammalian subject. Such compositions comprise, for instance, a media additive or a therapeutically effective amount of an inhibitor or enhancer compound to be identified by an assay of the invention and a pharmaceutically acceptable carrier or  
25 excipient. Such carriers may include, but are not limited to, saline, buffered saline, dextrose, water, ethanol, surfactants, such as glycerol, excipients such as lactose and combinations thereof. The formulation can be chosen by one of ordinary skill in the art to suit the mode of administration. The chosen route of administration will be influenced by the predominant tissue or organ location of the FATP whose function is  
30 to be inhibited or enhanced. For example, for affecting the function of FATP4, a preferred administration can be oral or through a tube inserted into the stomach (e.g.,

direct stomach tube or nasopharyngeal tube), or through other means to accomplish delivery to the small intestine. The invention further relates to diagnostic and pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention.

5           Compounds of the invention which are FATPs, FATP fusion proteins, FATP mimetics, FATP gene-specific antisense poly- or oligonucleotides, inhibitors or enhancers of a FATP may be employed alone or in conjunction with other compounds, such as therapeutic compounds. The pharmaceutical compositions may be administered in any effective, convenient manner, including administration by  
10   topical, oral, anal, vaginal, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal, transdermal or intradermal routes, among others. In therapy or as a prophylactic, the active agent may be administered to an individual as an injectable composition, for example as a sterile aqueous dispersion, preferably isotonic.

          Alternatively, the composition may be formulated for topical application, for  
15   example, in the form of ointments, creams, lotions, eye ointments, eye drops, ear drops, mouthwash, impregnated dressings and sutures and aerosols, and may contain appropriate conventional additives, including, for example, preservatives, solvents to assist drug penetration, and emollients in ointments and creams. Such topical formulations may also contain compatible conventional carriers, for example cream  
20   or ointment bases, and ethanol or oleyl alcohol for lotions.

          In addition, the amount of the compound will vary depending on the size, age, body weight, general health, sex, and diet of the host, and the time of administration, the biological half-life of the compound, and the particular characteristics and symptoms of the disorder to be treated. Adjustment and manipulation of established  
25   dose ranges are well within the ability of those of skill in the art.

          A further aspect of the invention is a method to identify a polymorphism, or the presence of an alternative or variant allele of a gene in the genome of an organism (of interest here, genes encoding FATPs). As used herein, polymorphism refers to the occurrence of two or more genetically determined alternative sequences or alleles  
30   in a population. A polymorphic locus may be as small as a base pair. Polymorphic markers include restriction fragment length polymorphisms, variable number of

tandem repeats (VNTR's), hypervariable regions, minisatellites, dinucleotide repeats, trinucleotide repeats, tetranucleotide repeats, simple sequence repeats, and insertion elements such as Alu. The first identified allelic form, or the most frequently occurring form can be arbitrarily designated as the reference (usually, "wildtype")  
5 form, and other allelic forms are designated as alternative (sometimes, "mutant" or "variant"). Diploid organisms may be homozygous or heterozygous for allelic forms.

An "allele" or "allelic sequence" is an alternative form of a gene which may result from at least one mutation in the nucleotide sequence. Alleles may result in altered mRNAs or polypeptides whose structure or function may or may not be  
10 altered. Any given gene may have none, one, or many allelic forms (polymorphism). Common mutational changes which give rise to alleles are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

15 Several different types of polymorphisms have been reported. A restriction fragment length polymorphism (RFLP) is a variation in DNA sequence that alters the length of a restriction fragment (Botstein *et al.*, *Am. J. Hum. Genet.* 32:314-331 (1980)). The restriction fragment length polymorphism may create or delete a restriction site, thus changing the length of the restriction fragment. RFLPs have  
20 been widely used in human and animal genetic analyses (see WO 90/13668; WO 90/11369; Donis-Keller, *Cell* 51:319-337 (1987); Lander *et al.*, *Genetics* 121:85-99 (1989)). When a heritable trait can be linked to a particular RFLP, the presence of the RFLP in an individual can be used to predict the likelihood that the individual will also exhibit the trait.

25 Other polymorphisms take the form of short tandem repeats (STRs) that include tandem di-, tri- and tetra-nucleotide repeated motifs. These tandem repeats are also referred to as variable number tandem repeat (VNTR) polymorphisms. VNTRs have been used in identity and paternity analysis (US 5,075,217; Armour *et al.*, *FEBS Lett.* 307:113-115 (1992); Horn *et al.*, WO 91/14003; Jeffreys, EP  
30 370,719), and in a large number of genetic mapping studies.

Other polymorphisms take the form of single nucleotide variations between individuals of the same species. Such polymorphisms are far more frequent than RFLPs, STRs (short tandem repeats) and VNTRs (variable number tandem repeats). Some single nucleotide polymorphisms occur in protein-coding sequences, in which  
5 case, one of the polymorphic forms may give rise to the expression of a defective or other variant protein and, potentially, a genetic disease. Other single nucleotide polymorphisms occur in noncoding regions. Some of these polymorphisms may also result in defective protein expression (e.g., as a result of defective splicing). Other single nucleotide polymorphisms have no phenotypic effects.

10 Many of the methods described below require amplification of DNA from target samples and purification of the amplified products. This can be accomplished by PCR, for instance. See generally, *PCR Technology, Principles and Applications for DNA Amplification* (ed. H.A. Erlich), Freeman Press, New York, NY, 1992; *PCR Protocols: A Guide to Methods and Applications* (eds. Innis, et al.), Academic Press,  
15 San Diego, CA, 1990; Mattila *et al.*, *Nucleic Acids Res.* 19:4967 (1991); Eckert *et al.*, *PCR Methods and Applications* 1:17 (1991); *PCR* (eds. McPherson *et al.*, IRS Press, Oxford); and US 4,683,202.

Other suitable amplification methods include the ligase chain reaction (LCR) (see Wu and Wallace, *Genomics* 4:560 (1989); Landegren *et al.*, *Science* 241:1077  
20 (1988)), transcription amplification (Kwoh *et al.*, *Proc. Natl. Acad. Sci. USA* 86:1173 (1989), self-sustained sequence replication (Guatelli *et al.*, *Proc. Natl. Acad. Sci. USA* 87:1874 (1990), and nucleic acid based sequence amplification (NASBA). The latter two amplification methods involve isothermal reactions based on isothermal transcription, which produce both single stranded RNA (ssRNA) and  
25 double stranded DNA (dsDNA) as the amplification products in a ratio of about 30 or 100 to 1, respectively.

Another aspect of the invention is a method for detecting a variant allele of a human FATP gene, comprising preparing amplified, purified FATP DNA from a reference human and amplified, purified, FATP DNA from a "test" human to be  
30 compared to the reference as having a variant allele, using the same or comparable amplification procedures, and determining whether the reference DNA and test DNA



differ in DNA sequence in the FATP gene, whether in a coding or a noncoding region, wherein, if the test DNA differs in sequence from the reference DNA, the test DNA comprises a variant allele of a human FATP gene. The following is a discussion of some of the methods by which it can be determined whether the  
5 reference FATP DNA and test FATP DNA differ in sequence.

Direct Sequencing. The direct analysis of the sequence of variant alleles of the present invention can be accomplished using either the dideoxy chain termination method or the Maxam and Gilbert method (see Sambrook *et al.*, *Molecular Cloning: A*  
10 *Laboratory Manual*, 2nd ed., Cold Spring Harbor Press, New York 1989; Zyskind *et al.*, *Recombinant DNA Laboratory Manual*, Acad. Press, 1988)).

Denaturing Gradient Gel Electrophoresis. Amplification products generated using the polymerase chain reaction can be analyzed by the use of denaturing gradient gel electrophoresis. Different alleles can be identified based on the different sequence-  
15 dependent strand dissociation properties and electrophoretic migration of DNA in solution (chapter 7 in Erlich, ed. *PCR Technology, Principles and Applications for DNA Amplification*, W.H. Freeman and Co., New York, 1992).

Single-strand Conformation Polymorphism Analysis. Alleles of target sequences can be differentiated using single-strand conformation polymorphism  
20 analysis, which identifies base differences by alteration in electrophoretic migration of single stranded PCR products, as described in Orita *et al.*, *Proc. Natl. Acad. Sci. USA* 86:2766-2770 (1989). Amplified PCR products can be generated as described above, and heated or otherwise denatured, to form single-stranded amplification products. Single-stranded nucleic acids may refold or form secondary structures  
25 which are partially dependent on the base sequence. The different electrophoretic mobilities of single-stranded amplification products can be related to base-sequence differences between alleles of target sequences.

Detection of Binding by Protein That Binds to Mismatches. Amplified DNA comprising the FATP gene or portion of the gene of interest from genomic DNA, for  
30 example, of a normal individual is prepared, using primers designed on the basis of the DNA sequences provided herein. Amplified DNA is also prepared, in a similar

manner, from genomic DNA of an individual to be tested for bearing a distinguishable allele. The primers used in PCR carry different labels, for example, primer 1 with biotin, and primer 2 with  $^{32}\text{P}$ . Unused primers are separated from the PCR products, and the products are quantitated. The heteroduplexes are used in a mismatch detection assay using immobilized mismatch binding protein (MutS) bound to nitrocellulose. The presence of biotin-labeled DNA wherein mismatched regions are bound to the nitrocellulose via MutS protein, is detected by visualizing the binding of streptavidin to biotin. See WO 95/12689. MutS protein has also been used in the detection of point mutations in a gel-mobility-shift assay (Lishanski, A. *et al.*, *Proc. Natl. Acad. Sci. USA* 91:2674-2678 (1994)).

Other methods, such as those described below, can be used to distinguish a FATP allele from a reference allele, once a particular allele has been characterized as to DNA sequence.

Allele-specific probes. The design and use of allele-specific probes for analyzing polymorphisms is described by e.g., Saiki *et al.*, *Nature* 324:163-166 (1986); Dattagupta, EP 235,726, Saiki, WO 89/11548. Allele-specific probes can be designed so that they hybridize to a segment of a target DNA from one individual but do not hybridize to the corresponding segment from another individual due to the presence of different polymorphic forms in the respective segments from the two individuals. Hybridization conditions should be sufficiently stringent that there is a significant difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the alleles. Some probes are designed to hybridize to a segment of target DNA such that the polymorphic site aligns with a central position (e.g., in a 15-mer at the 7 position; in a 16-mer, at either the 8 or 9 position) of the probe. This design of probe achieves good discrimination in hybridization between different allelic forms.

Allele-specific probes are often used in pairs, one member of a pair showing a perfect match to a reference form of a target sequence and the other member showing a perfect match to a variant form. Several pairs of probes can then be immobilized on the same support for simultaneous analysis of multiple polymorphisms within the same target sequence.

Allele-specific Primers. An allele-specific primer hybridizes to a site on target DNA overlapping a polymorphism, and only primes amplification of an allelic form to which the primer exhibits perfect complementarity. See Gibbs, *Nucleic Acid Res.* 17:2427-2448 (1989). This primer is used in conjunction with a second primer  
5 which hybridizes at a distal site. Amplification proceeds from the two primers, resulting in a detectable product which indicates the particular allelic form is present. A control is usually performed with a second pair of primers, one of which shows a single base mismatch at the polymorphic site and the other of which exhibits perfect complementarity to a distal site. The single-base mismatch prevents amplification  
10 and no detectable product is formed. The method works best when the mismatch is included in the 3'-most position of the oligonucleotide aligned with the polymorphism because this position is most destabilizing to elongation from the primer (see, e.g., WO 93/22456).

Gene Chips. Allelic variants can also be identified by hybridization to nucleic  
15 acids immobilized on solid supports (gene chips), as described, for example, in WO 95/11995 and U.S. Patent No. 5,143,854, both of which are incorporated herein by reference. WO 95/11995 describes subarrays that are optimized for detection of a characterized variant allele. Such a subarray contains probes designed to be complementary to a second reference sequence, which is an allelic variant of the first  
20 reference sequence.

The present method is illustrated by the following examples, which are not intended to be limiting in any way.

## EXAMPLES

### Materials and Methods

25 The following Materials and Methods were used in the work described in Examples 1-5.

Sequence Alignment of FATP Clones. The DNA sequence for mouse FATP1 was obtained from the National Center for Biotechnology Information nonredundant database. cDNAs for mmFATP2, 3, 4, and 5 were obtained by screening mouse

expression libraries (purchased from GIBCO/BRL, Rockville, MD) with probes derived from the cloned expressed sequence tags (ESTs) (Research Genetics, Huntsville, AL). Full-length clones were obtained for mmFATP2 and 5 and partial sequences for mmFATP3 and 4. The sequences described herein have been  
5 deposited in the GenBank database (Accession Nos. FATP2, AF072760; FATP3, AF072759; FATP4, AF072758; FATP5, AF072757).

Neither FATP2 nor FATP5 contains an in-frame stop codon upstream of the putative initiator methionine; initiator methionines were assigned by homology with that in mmFATP1 and by the presence of a signal sequence immediately after it. The  
10 *Mycobacterium tuberculosis*, *Caenorhabditis elegans*, and *Saccharomyces cerevisiae* sequences were present in the dbEST database as part of the sequencing projects for these organisms. Sequences were aligned utilizing a ClustalX algorithm and the resulting alignment exported to SeqVu. Homologous amino acid substitutions are boxed in Figure 1 and were determined using the Dayhoff 250 method with a 50%  
15 homology cutoff.

Cell Transfection and LCFA Uptake. COS cells were cotransfected using the DEAE-dextran method with the mammalian expression vector pCDNA 3.1 (Invitrogen, Carlsbad, CA) expressing the gene for CD2 (pCDNA-CD2) in combination with either a pCDNA 3.1 or pCMVSPORT2 (GIBCO/BRL, Rockville,  
20 MD) expression vector containing one of the murine or nematode *FATP* genes (*pCDNA-mmFATP1*, *pCDNA-FATP2*, *pCMVSPORT-FATP5*, *pCDNA-ceFATPb*). Two days after transfection, cells were assayed for CD2 expression with a phycoerythrin-coupled anti-CD2(PE-CD2) monoclonal antibody (PharMingen, Franklin Lakes, NJ), and fatty acid uptake was assayed with a BODIPY-labeled fatty  
25 acid analogue (Molecular Probes). Briefly, cells were washed twice with PBS (phosphate buffered saline) and stained with PE-CD2 at 4°C for 30 min in PBS containing 10% fetal calf serum. They were then washed three times with PBS/fetal calf serum for 5 min followed by an incubation for 2 min at 37°C in fatty acid uptake solution, which contained 0.1 µM BODIPY-FA and 0.1% fatty acid-free BSA  
30 (bovine serum albumin) in PBS (Schaffer, J.E. & Lodish, H.F. (1994) *Cell* 79:427-436). After 2 min, the cells were washed four times with ice-cold PBS/0.1% BSA.

The cells were then removed from the plates with PBS containing 5 mM EDTA and resuspended in PBS containing 10% fetal calf serum and 10 mM EDTA. PE-CD2 and BODIPY-FA fluorescence were measured using a FACScan (Becton Dickinson, Franklin Lakes, NJ). COS cells were gated on forward scatter (FSC) and side scatter (SS). Cells exhibiting more than 300 CD2 fluorescence units (dsim) representing 15% of all cells were deemed CD2 positive and their BODIPY-FA fluorescence was quantitated.

*E. coli*-Based LCFA Uptake Assay. The full-length coding region of mtFATP and a control protein, the mammalian transcription factor TFE3, were subcloned into the inducible, prokaryotic expression vector pET (Novagen, Madison, WI). Expression was induced with 1 mM isopropyl  $\beta$ -D-thiogalactoside (IPTG) for 1 hour, or cells were left uninduced. Cells were washed in PBS/0.1% BSA and resuspended in 1 ml PBS/0.1% BSA containing 0.1  $\mu$ M [ $^3$ H]palmitate (NEN) at 37°C. Uptake was stopped after the indicated incubation time by transferring the cells onto filter paper using a cell harvester (Brandel, Bethesda, MD). Filters were washed extensively with ice-cold PBS/0.1% BSA, and [ $^3$ H]palmitate was quantitated by scintillation counting.

Northern Blots. Northern blot analysis of murine FATP expression was done using poly(A) mRNA blots (Clontech, Palo Alto, CA). Probes of each of the FATPs were derived from the 3' untranslated regions of each gene and were <60% identical in sequence. Probes were labeled by random priming (Boehringer Mannheim, Indianapolis, IN) and hybridized at 65°C. Blots were extensively washed in 0.2% SSC/0.1% SDS at 65°C.

Generation of Phylogenetic Trees. Complete and partial sequences for *FATP* genes from human, rat, mouse, puffer fish, *Drosophila melanogaster*, *C. elegans*, *S. cerevisiae*, and *M. tuberculosis* were aligned using ClustalX. A homologous region of 48 amino acids (residues 472-519 in mmFATP1) from all of the genes was used to determine phylogenetic relationship within ClustalX. Based on these data a phylogenetic tree was generated using Tree View PPC (Figure 5).

Nomenclature. It is proposed that the *FATP* genes be given a species specific prefix (mm, *Mus musculus*; hs, *Homo sapiens*; mt, *M. tuberculosis*; dm, *D.*

*melanogaster*; ce, *C. elegans*, sc, *S. cerevisiae*) and numbered such that mammalian homologues in different species share the same number but differ in their prefix. Since the two *C. elegans* genes cannot be paired with a specific human or mouse FATP, they have been designated *ceFATPa* and *ceFATPb*.

5    Example 1: Identification of Novel Mammalian FATPs

The National Center for Biotechnology Information EST database was screened, using the mouse FATP protein sequence (mmFATP1), to identify novel FATPs. This strategy led to the identification of more than 50 murine EST sequences which could be assembled into five distinct contiguous DNA sequences (contigs).

- 10    One contig was identical to the previously cloned FATP, which has been renamed FATP1. Another, which has been renamed FATP2, is the murine homologue of a rat gene previously identified by others as a very long chain acyl-CoA synthase (Uchiyama, A., Aoyama, T., Kamijo, K., Uchida, Y., Kondo, N., Orii, T. & Hashimoto, T. (1996) *J. Biol. Chem.* 271:30360-30365). The other three contigs
- 15    represented novel genes (*FATP3*, 4, and 5). Full-length clones for *FATP2* and *FATP5* and nearly complete sequences for *FATP3* and 4 (Figure 1) were obtained by screening cDNA libraries made from mouse day 10.5 embryos and adult liver. Also identified were human homologues for each of the murine genes in the EST database. A sixth human gene was also identified; whether this gene is also present in the
- 20    mouse will require additional studies. Map positions are given in Tables 2 and 3.

The genetic loci for all of the human genes, with the exception of *FATP5* which was already mapped as an unknown EST, were determined using the radiation hybrid

- panels. The map positions given below show the distance (in centiRays) from the
- 25    closest framework marker. As a guideline, there are approximately 300 kb/cR.

Table 2. Mapping Data for Human Genes

5	hsFATP1	Chromosome Chr19 places 13.35 cR from WI-6344 (lod>3.0)
	hsFATP2	Chromosome Chr15 places 4.92 cR from D15S126 (lod>3.0)
	hsFATP3	Chromosome Chr1 places 13.24 cR from WI-2862 (lod>3.0)
	hsFATP4	Chromosome Chr9 places 7.80 cR from WI-9685 (lod>3.0)
	hsFATP5	unknown EST previously mapped to near D19S418
10	hsFATP6	Chromosome Chr5 places 1.41 cR from WI-4907 (lod>3.0)

The mouse map is an internal backcross panel consisting of 188 mouse backcross DNA's plus 4 controls (B6, Spretus, F1, Water). The backcross was constructed by crossing B6 by Spretus animals and then crossing those F1's back to B6. Mapping is accomplished by taking advantage of recombinational events during meiosis, and the use of PCR primers to detect the differences (by size or re-annealing events) at any given locus between the B6 and Spretus allele.

For the purposes of mapping, a novel set of primers (gene of interest) is used to amplify from all 188 DNA's and then typed as being a B6 ("B") or a Spretus ("S"). This string of B's and S's is entered into the Map Manager program, which does a best fit calculation by comparing the string of 188 typings from the gene of interest to all loci already extant in the panel, for all 20 chromosomes. The gene of interest is then assigned to a particular area on a particular chromosome according to a number of parameters, including the minimalization of double cross-overs, and the highest LOD scores. Indicated in Table 3 are distances to the closest markers on either side of the FATP locus.

Table 3. Mapping Data for Mouse Genes

	mmFATP1	Chromosome 8 places 2.82 cM from D8Mit132 (lod 43.4) and 1.81 cM from D8Mit74 (lod 43.5)
5	mmFATP2	Chromosome 2 places 1.29 cM from D2Mit258 (lod 47.9) and 1.75 cM from D2NDS3 (lod 44.9)
	mmFATP3	Chromosome 3 places 2.54 cM from D3Mit22 (lod 29.5) and 19.62 cM from D3Mit42 (lod 13.6)
10	mmFATP4	Chromosome 2 places 13.78 cM from D2Mit1 (lod 22.9) and 3.85 cM from D2Mit65 (lod 41.9)
	mmFATP5	Chromosome 7 places 7.28 cM proximal of D7Mit21 (lod 28.3)
15		

## Example 2: Assessment of Function

The ability of the newly identified mouse genes to function as fatty acid transporters was assessed using a fluorescence-activated cell sorting-based assay. COS cells were transiently cotransfected with expression vectors encoding the cell surface protein CD2 and either mmFATP1, mmFATP2, or mmFATP5, respectively. Two days after transfection, COS cells were stained with an antibody to CD2 and then incubated with a BODIPY-labeled fatty acid [BODIPY-FA, (Schaffer, J.E. & Lodish, H.F. (1994) *Cell* 79:427-436)]. The cells were then washed extensively, lifted off the dish, and analyzed by fluorescence-activated cell sorting. As judged by the number of CD2-positive cells, the transfection efficiency was approximately 20-30%. Fatty acid uptake was quantitated in the transiently transfected COS cells by measuring the BODIPY-FA fluorescence of the CD2-positive cells. Expression of CD2 had no effect on fatty acid uptake as shown by the finding that COS cells expressing only the transfected CD2 cDNA (CD2-positive) had the same low level of



BODIPY-FA uptake as did untransfected (CD2-negative) control cells (Figure 2A, control). In COS cells cotransfected with CD2 and mmFATP1, mmFATP2, or mmFATP5, uptake of BODIPY-FA by the transfected (CD2-positive) cells was increased between 15- to 90-fold over control (CD2 cDNA only) cells (Figures 2A-  
5 2D).

### Example 3: Expression Patterns of Murine FATPs

Expression patterns of members of the murine *FATP* gene family were characterized by Northern blot analysis; to avoid cross-hybridization, the probes used were from the 3' untranslated region of these genes, which are less than 60% identical  
10 in sequence. The expression pattern of FATP1 agrees with that previously found (Schaffer, J.E. & Lodish, H.F. (1994) *Cell* 79:427-436). Here, expression was seen primarily in heart and kidney. FATP2 is expressed almost exclusively in liver and kidney, which corresponds to the reported tissue distribution of the rat homologue [very long chain acyl-CoA (VLACS)] as assessed by Western blotting (Uchiyama, A.,  
15 Aoyama, T., Kamijo, K., Uchida, Y., Kondo, N., Orii, T. & Hashimoto, T. (1996) *J. Biol. Chem.* 271:30360-30365). FATP3 is present in lung, liver, and testis. FATP5 is expressed only in liver and cannot be detected in other tissues even when the blot is overexposed. The human homologue of FATP5 is also liver specific and is not expressed in a wide array of other tissues tested, including fetal liver.

### 20 Example 4: FATPs Are Evolutionarily Conserved

The EST database was searched, using sequences conserved among the five murine *FATP* genes, for *FATP* genes in other organisms. Two homologues were found in *C. elegans* and one in *M. tuberculosis*. One of the *C. elegans* genes was cloned from a cDNA library and expressed in COS cells, as described for the murine  
25 *FATPs*. Overexpression of the nematode *FATP* resulted in a 15-fold increase of BODIPY-FA uptake compared with control cells (Figure 3). The mycobacterial *FATP* gene was isolated from a phage library and assessed for its ability to facilitate fatty acid uptake. *E. coli* transformed with a prokaryotic, isopropyl  $\beta$ -D-thiogalactoside-inducible expression vector containing the mycobacterial *FATP* gene

demonstrated a significant increase in the rate of [ $^3\text{H}$ ]palmitate uptake after induction, compared with uninduced bacteria or *E. coli* transformed with a control protein (Figure 4). Novel *FATP* genes were also identified in *F. rubripes* (puffer fish) and *D. melanogaster*.

#### 5 Example 5: Phylogenetic Tree of FATPs

Faergeman *et al.* (Faergeman, N.J., DiRusso, C.C., Elberger, A., Knudsen, J. & Black, P. N. (1997) *J. Biol. Chem.* 272:8531-8538) identified three regions of very strong conservation between the *scFATP* and *mmFATP1* genes. The sequences of the FATPs were compared over a 311-amino acid FATP "signature sequence" which  
10 includes these conserved regions corresponding to amino acids 246-557 in *mmFATP1* (underlined in Figure 1). When compared with the National Center for Biotechnology Information nonredundant database, only one region of the "FATP signature sequence" shows significant homology to other proteins. This small stretch  
15 of amino acids (underlined in Fig. 1) is an AMP-binding motif found in a multitude of other proteins, such as acyl-CoA synthase, several CoA lipases, and gramicidin S synthetase component II (Schaffer, J.E. & Lodish, H.F. (1994) *Cell* 79:427-436). The relevance of this motif to fatty acid transport is unclear. Other highly conserved regions among the FATPs, including long stretches of amino acids >90% identical from mycobacteria to humans, are not found in any other class of proteins. A 48-  
20 amino acid segment of the FATP signature sequence was used to construct a phylogenetic tree (Figure 5). Each of the human and mouse genes form their own branch; *hsFATP6*, which as yet has no murine homologue, is most closely related to *hsFATP3* and *mmFATP3*. As expected, *mVLACS* is closer in sequence to *mmFATP2* than to *hsFATP2*. The *FATP* genes of invertebrates i.e., *C. elegans* and  
25 *D. melanogaster*, are most closely related to each other. Surprisingly, the mycobacterial gene is more closely related to the human and mouse *FATP5* genes than to the FATPs of any of the lower organisms. Whether this reflects coevolution of the mycobacterial and human genes awaits further study.

## Materials and Methods

The following materials and methods were used in the work described in Examples 6-10.

### Isolation of full-length human FATP1 and 4

- 5 Full-length clones encoding human FATP1 and human FATP4 were identified by searching databases for sequences similar to murine FATP1-5 coding regions using the BlastX algorithm (Altschul *et al.*, *J. Mol. Biol.* 215: 403-410, 1990).

A concatamer of nucleotide sequences comprising the coding sequences of mmFATP1 (Genbank Accession U15976), mmFATP2, mmFATP3 (SEQ ID NO:6),  
10 mmFATP4 (SEQ ID NO:8) and mmFATP5 (SEQ ID NO:10) was used to search the Millennium database using the BLASTX algorithm. Sequences with a score >150 were evaluated for whether they represented known FATP coding sequences.

Human clones with similarity to the 5' end of murine FATP sequences were sequenced completely. Clones encoding full-length human FATP1 were obtained  
15 from a heart cDNA library constructed in the mammalian expression vector pMET7 (Tartaglia *et al.*, *Cell*, 83: 1263-1271, 1995). Clones encoding full-length human FATP4 were obtained from a spleen cDNA library constructed in the mammalian expression vector pMET7.

### Isolation of full-length human FATP6

- 20 Several clones encoding human FATP6 were identified by searching public databases as described above. Five clones were analyzed further by restriction digestion and DNA sequencing. One of these clones (Genbank Accession # AA412064) appeared to be full-length and its entire insert was sequenced.

### DNA Sequence Analysis

- 25 Sequences were aligned with the DNASTar program using the Clustal method. Hydrophobicity plots were generated with DNA Strider using the Kyte Doolittle method.

### In situ hybridization

Tissues were collected from 8 week old C57/B16 mice. Tissues were fresh frozen, cut on a cryostat at 10  $\mu$ m thickness and mounted on Superfrost Plus slides (VWR). Sections were air dried for 20 minutes and then incubated with ice cold 4% paraformaldehyde (PFA)/phosphate buffered saline (PBS) for 10 minutes. Slides were washed 2 times 5 minutes with PBS, incubated with 0.25% acetic anhydride/1 M triethanolamine for 10 minutes, washed with PBS for 5 minutes and dehydrated with 70%, 80%, 95% and 100% ethanol for 1 minute each. Sections were incubated with chloroform for 5 minutes. Hybridizations were performed with  $^{35}$ S-radiolabeled (5x10<sup>7</sup> cpm/ml) cRNA probes generated from the 3' untranslated regions of mouse FATPs by PCR followed by *in vitro* transcription in the presence of 50% formamide, 10% dextran sulfate, 1x Denhardt's solution, 600 mM NaCl, 10 mM DTT, 0.25% SDS and 10  $\mu$ g/ml tRNA for 18 hours at 55°C. After hybridization, slides were washed with 10 mM Tris-HCl pH 7.6, 500 mM NaCl, 1 mM EDTA (TNE) for 10 minutes, incubated in 40  $\mu$ g/ml RNase A in TNE at 37°C for 30 minutes, washed in TNE for 10 minutes, incubated once in 2x SSC at 60°C for 1 hour, once in 0.2x SSC at 60°C for 1 hour, once in 0.2x SSC at 65°C for 1 hour and dehydrated with 50%, 70%, 80%, 90% and 100% ethanol. Localization of mRNA transcripts was detected by dipping slides in Kodak NBT-2 photoemulsion and exposing for 7 days at 4°C, followed by development with Kodak Dektol developer. Slides were counter stained with haematoxylin and eosin and photographed. Controls for the in situ hybridization experiments include the use of a sense probe which showed no signal above background in all cases.

### Northern Blotting

Human mRNA blots were obtained from Invitrogen or Clontech. PCR fragments from the 3' untranslated regions of human FATPs were used as probes. Blots were probed with  $^{32}$ P-labeled DNA probes using the Rapid-Hyb buffer (Amersham, Buckinghamshire, UK) according to the manufacturer's instructions.

Cell transfection and LCFA uptake. COS cells were cotransfected, using lipofectamine (GIBCO BRL, Rockville, MD) according to the manufacturer's

instructions, with the mammalian expression vector pCDNA3.1 (Invitrogen, Carlsbad, CA) expressing the gene for CD2 in combination with a pMET7 expression vector (Tartaglia *et al.*, *Cell*, 83:1263-1271, 1995) containing hsFATP1 (pMET7-hsFATP1) or hsFATP4 (pMET7-hsFATP4) or pMET7 alone. Two days after  
5 transfection, cells were assayed for CD2 expression with a phycoerythrin-coupled anti-CD2 (PE-CD2) monoclonal antibody (PharMingen, Franklin Lakes, NJ), and fatty acid uptake was assayed with a BODIPY-labeled fatty acid analog (Molecular Probes) as described above.

#### Example 6: Determination of Expression of mmFATPs

10 mmFATP4, and to lesser extent mmFATP2, are expressed at high levels in the brush border layer of the small intestine.

Cell transfection and LCFA uptake. COS cells were cotransfected, using lipofectamine (GIBCO BRL, Rockville, MD) according to the manufacturer's instructions, with the mammalian expression vector pCDNA3.1 (Invitrogen, Carlsbad, CA) expressing the gene for CD2 in combination with a pMET7 expression  
15 vector (Tartaglia *et al.*, *Cell*, 83:1263-1271, 1995) containing hsFATP1 (pMET7-hsFATP1) or hsFATP4 (pMET7-hsFATP4) or pMET7 alone. Two days after transfection, cells were assayed for CD2 expression with a phycoerythrin-coupled anti-CD2 (PE-CD2) monoclonal antibody (PharMingen, Franklin Lakes, NJ), and  
20 fatty acid uptake was assayed with a BODIPY-labeled fatty acid analog (Molecular Probes) as described above.

Absorption of dietary fat requires transport of free fatty acids across the apical membrane of epithelial cells in the small intestine. Previous studies suggested that this transport is protein-mediated; however, the transport protein had not yet been  
25 identified. In situ hybridization was performed on each of the three regions of the small intestine -- duodenum, jejunum and ileum -- as well as the colon, using probes from the 3' untranslated regions of mmFATP1, mmFATP2, mmFATP3, mmFATP4 and mmFATP5, to determine whether any of the mouse FATPs are expressed in the small intestine. It was expected that a protein involved in fatty acid absorption would  
30 be expressed in the epithelial cells of the small intestine, but absent from the colon.

Expression of mmFATPs in the jejunum was identical to that in the ileum in all cases. High levels of mmFATP4 mRNA were present in the epithelial cells of the jejunum and ileum, and lower, but significant, amounts were detected in the epithelial cells of the duodenum. Significantly, FATP4 mRNA was absent from other cell  
5 types of the small intestine and no FATP4 mRNA could be detected in any of the cells of the colon. FATP2 mRNA was present in the epithelial cells of the duodenum at a level similar to that of FATP4, but was present at lower levels in the jejunum and ileum. No signals above background were detected for mmFATP1, mmFATP3 and mmFATP5 in any of the intestinal tissues. mmFATP3 and FATP5 were clearly  
10 detectable by in situ hybridization in adult liver and mmFATP1 could be detected in a variety of tissues on a whole embryo in situ, indicating that the FATP1, 3, and 5 probes were working.

mmFATP4 expression is predominant in the small intestine compared to the other organs of the mouse embryo. In the small intestine, FATP4 expression is  
15 limited to differentiated enterocytes, while no signal is detected in the connective tissue or the undifferentiated epithelial cells in the crypts. Differentiated enterocytes are known to be the cells that mediate the uptake of fatty acids. FATP4 is specifically and strongly expressed in the epithelial cells of adult murine duodenum and ileum but not colon. Other FATPs, such as FATP5, are not expressed in the small intestine.  
20 Thus, FATP4 is the major FATP in the mouse small intestine. Given its high level of expression, it is likely that FATP4, and to a lesser extent FATP2, play an important role in the absorption of fatty acids.

mmFATP2, and mmFATP5 are expressed in hepatocytes

Northern analysis of mmFATP2, mmFATP3, mmFATP4 and mmFATP5  
25 showed expression in the liver. To determine whether these proteins are present in hepatocytes or other cells types present in liver homogenates, in situ hybridizations were performed. mmFATP2, and mmFATP5 mRNA was clearly present in hepatocytes, and was not concentrated in other cell types such as endothelial cells or macrophages. No signal above background was detected for mmFATP1 in any of the  
30 cell types in the liver, consistent with the results of the Northern blotting.

Example 7: Isolation and Sequence Analysis of Full-length Human FATP1 and Full-length Human FATP4

To identify human cDNA clones encoding FATP family members, Millennium databases were searched for sequences similar to murine FATP1-5 coding regions. Two clones were analyzed in detail; inspection of the entire DNA sequence of these two clones showed that they encode the human orthologs of mmFATP1 and mm FATP4, respectively. These two clones were designated hsFATP1 and hsFATP4, and their DNA and predicted protein sequences are shown in Figures 44A-44C and 45, and 50A-50C and 51. hsFATP1 is predicted to encode a 646 amino acid, 71 kD protein with multiple membrane-spanning domains (Figure 28A). HsFATP4 is predicted to encode a 643 amino acid, 72 kD protein with multiple membrane spanning domains (See Figure 29A). A comparison of the DNA sequences of mouse and human FATP1 and mouse and human FATP4 (Figures 30A-30B and 31A-31B) shows that the mouse and human orthologs are 85% (FATP1) and 87% (FATP4) identical to each other within the coding sequences given in these figures. At the amino acid level, hsFATP1 and hsFATP4 are ~90% identical to their respective mouse orthologs within the coding region shown in these figures (Figures 32 and 33). The sequence identities between mouse and human FATP1 and FATP4 are considerably higher than the ones observed between different FATP family members within one species (~40%-60%) and are present in the N-terminal part of the protein, a region that is poorly conserved between different FATP family members. This high degree of sequence conservation clearly demonstrates that the newly identified human FATPs are orthologs of mouse FATP1 and FATP4 rather than novel FATP family members.

Table 4 is an identity/similarity matrix comparing the amino acid sequences of FATP1 and 4 from human and mouse. This shows that the gene whose sequence is shown in Figure 43A is indeed human FATP4, since it is 91% identical with the murine FATP4 but only 62% identical with the closest related human FATP, which is FATP1.

Table 4				
Identity/Similarity Matrix				
	hsFATP4	mmFATP4	hsFATP1	mmFATP1
hsFATP4	---	93.2	72.3	72.0
mmFATP4	91.0	---	71.2	71.1
hsFATP1	61.9	61.0	---	92.4
mmFATP1	60.7	59.6	89.5	---

5

#### Example 8: Isolation and Sequence Analysis of Full-length Human FATP6

A search of EST databases identified a set of overlapping human sequences that were similar to FATPs, but did not have a clear mouse ortholog. One of these EST clones was found to encode a full-length cDNA. The entire insert of this clone was sequenced and designated hsFATP6. The DNA and predicted protein sequences of hsFATP6 are shown in Figures 54A-54C and 55. HsFATP6 is predicted to encode a 619 amino acid, 70 kD protein with multiple membrane-spanning domains (Figure 35A). A comparison of the amino acid sequences of hsFATP6 with other human FATPs shows about 37% identity to either hsFATP1 or hsFATP4 (Figure 36). This degree of sequence identity is similar to what is observed between different mouse FATPs. The phylogenetic analysis described above clearly demonstrates that hsFATP6 is a member of the FATP family, but not an ortholog of any of the mouse FATPs. Comparisons were done with "ALIGN" (E. Myers and W. Miller, "Optimal Alignments in Linear Space," *CABIOS* 4:11-17 (1988) using standard settings.

#### Example 9: Tissue Distribution of Human FATPs

The tissue distribution of human FATPs was assessed by Northern blotting. Human FATP3 was expressed in a large variety of tissues. In contrast, human FATP5 was present at high levels in the liver, but was undetectable in all other tissues examined. Thus, both hsFATP3 and hsFATP5 recapitulate the expression

25



pattern of their mouse orthologs (see above). HsFATP6 is a novel FATP with no mouse ortholog as yet. Northern blotting shows that hsFATP6 is expressed at high levels in the heart, but is undetectable in other tissues, including skeletal and smooth muscle. This tissue distribution suggests that human FATP6 performs an important  
5 role in energy metabolism in the heart; blocking FATP6-mediated fatty acid transport may therefore be beneficial for a number of heart diseases, e.g., ischemic heart disease.

To identify the major FATP expressed in the human small intestine, Northern blotting was performed on a blot containing mRNA from human stomach, jejunum,  
10 ileum, colon, rectum and lung. hsFATP5 and hsFATP6 were undetectable in any of these tissues. FATP5 is only expressed in liver and FATP6 only in heart. hsFATP2 was weakly expressed in the colon, and an even weaker signal was detectable in jejunum, ileum and lung lanes. hsFATP3 was expressed well in the lung, but was only weakly expressed in the other tissues tested. Importantly, no difference was seen  
15 in the expression of hsFATP3 between small intestine and stomach or colon, suggesting that the expression observed is not related to fatty acid absorption in the small intestine. hsFATP4 was clearly expressed in both jejunum and ileum; expression was significantly lower in the colon and was absent in the stomach. This expression pattern is consistent with a major role for FATP4 in absorption of fatty  
20 acids in the human gut.

#### Example 10: Expression of hsFATP1 and hsFATP4 Promotes Transport of Fatty Acids

COS cells were cotransfected using lipofectamine with the mammalian expression vector pCDNA-CD2 in combination with one of the FATP-containing  
25 expression vectors (pMET7-hsFATP1 or pMET7-hsFATP4) or an insertless expression vector (pMET7, control) as described in Materials and Methods for Examples 6-10. COS cells were gated on forward scatter and side scatter. Cells exhibiting more than 400 CD2 fluorescence units representing ~30% of all cells were deemed CD2-positive. The percent of CD2-positive cells exhibiting a BODIPY-  
30 fluorescence of >300 is plotted for the three different vectors tested (Figure 37).

#### Example 11: Stable Expression of Human FATP4 in 293 Cells

Stable cell lines were generated as follows. A DNA fragment containing the entire hsFATP4 coding sequence as well as 100 nucleotides of 5' and 50 nucleotides of 3' untranslated region was inserted into the vector pIRES-neo (Clontech, Palo Alto, CA) using standard cloning techniques. The resulting construct or a vector control (pIRES-neo) was transfected into 293 cells using the lipofectamine method (Gibco BRL, Rockville, MD) according to the manufacturer's directions. Cells that had taken up the DNA were selected with 1 mg/ml G418 (Gibco BRL, Rockville, MD). Single colonies were picked 1 to 2 weeks after transfection and grown in medium containing 0.8 mg/ml G418. Colonies were screened for the ability to take up fatty acids by measuring uptake of a fluorescently labeled fatty acid (BODIPY-FA). About 40 colonies transfected with the pIRES-neo containing FATP4 and ~20 colonies transfected with pIRES-neo control were analyzed. All 20 of the vector control clones showed amounts of BODIPY-FA uptake similar to each other and to untransfected 293 cells. In contrast, among the 40 FATP4 transfected clones, 3 had a 5- to 10-fold increased BODIPY-FA uptake compared to any of the vector controls, and a large number (~20) showed an approximately two-fold increase in BODIPY-FA levels. This distribution is consistent with FATP4 conferring increased fatty acid uptake in these cells. One of the cell lines with the highest amount of BODIPY-FA uptake was selected to be used for measuring uptake of tritiated fatty acid.

The uptake of tritiated oleate over time by either FATP4 expressing or control cells was assayed over time. Expression of FATP4 increases the rate of fatty acid uptake by over 3-fold, demonstrating that FATP4 is, like the other FATPs, a functional fatty acid transporter (Figure 38).

#### Example 12: Immuno-staining with FATP4-Specific Antiserum

A polyclonal antiserum against the C-terminus of mmFATP4 was raised using a GST-fusion protein having mmFATP4-specific amino acid sequence 552-643 (AVASP...GEEKL). In western blot experiments, the purified antibody reacted strongly with a synthetic peptide matching the C-terminus of mmFATP4, but not with a corresponding region of mmFATP2, mmFATP3, or mmFATP5. The mmFATP4

specific polyclonal antiserum detects, in western blot experiments with enterocyte lysates from 3 different mice, a ~70 kDa protein, which is in accordance with mmFATP4's predicted molecular weight of 72 kDa. The binding is specific for mmFATP4, since it can be completely abolished by preincubation of the antiserum  
5 with the GST-fusion peptide used to raise the antibody.

Immunofluorescence experiments were performed using the anti-mmFATP4 antiserum on fresh frozen sections of murine small intestine. The antibody binding demonstrates strong expression of mmFATP4 in enterocytes, confirming the results of the in situ hybridization experiments. At higher magnifications it is apparent that  
10 mmFATP4 is expressed at the apical side of the enterocyte, indicating that the transporter is present in the brush border membrane, which is known to mediate the uptake of fatty acids from the intestinal lumen.

Immuno-electron microscopy studies were performed on fresh frozen murine intestinal cells. The gold particles used, appearing as black specks on the electron  
15 micrographs, indicate the subcellular localization of mmFATP4 to be on the microvilli of the enterocyte. It can be seen from electron micrographs that mmFATP4 is localized exclusively in membranes, preferentially the apical plasma membrane, confirming that it is indeed a membrane protein.

#### Methods for Immunofluorescence and Immunogold Electron Microscopy

20 Unfixed mouse small intestine was washed with Hank's buffered salt solution containing 1 mM EDTA, infused with 2.3 M sucrose solution, and embedded in O.C.T., 4583 compound. The material was thick sectioned (15  $\mu$ M - 40  $\mu$ M). The sections were washed in PBS containing 1% BSA and 0.075% glycine to block non-specific binding. Primary and secondary antibodies were diluted in PBS with 10%  
25 FCS and incubated for 1h. The sections were mounted in 90% glycerol/PBS containing 1 mg/ml paraphenylenediamine, and examined with a Bio-Rad MRC 600 confocal, mounted on a Zeiss Axioscop.

For the immunogold labeling, the tissue was fixed with 2% paraformaldehyde in PBS for 10 minutes, after which it was cryoprotected by infiltration with 2.3 M  
30 sucrose in 0.1 M phosphate buffer (pH 7.4) containing 20% polyvinylpyrrolidone,

and then mounted on aluminum cryo nails and frozen in liquid nitrogen (Tokuyasu, K.T., *J. Microscop.* 143:139-149, 1986). Ultrathin sections were collected on carbon/formvar-coated nickel grids. The primary antibody (anti-FATP4) was diluted in 10% FCS in PBS and incubated overnight at 4° C, followed by donkey anti-rabbit  
 5 IgG-gold (12 nm) (Jackson Labs) for 1h. The sections were stained in 2% neutral uranyl acetate (20 minutes) and absorption stained with 2% uranyl acetate in 0.2% methylcellulose containing 3.2% polyvinyl alcohol. The sections were examined with a Philips EM 410 electron microscope.

Example 13: Inhibition of Fatty Acid Uptake Specific to FATP4 Demonstrated in  
 10 Isolated Mouse Enterocytes

Phosphorothioate derivatives of the following oligonucleotides were synthesized:

	FATP4-AS2	CCCCCACCAGAGAGGCTCC (SEQ ID NO:103)
	FATP4-AS2MM	CCACCCCCGGAAAGCCTGC (SEQ ID NO:104)
15	FATP4-S2	GGAGCCTCTCTGGTGGGGG (SEQ ID NO:105)

FATP4 AS2 is the antisense oligo; it is designed to be complementary to the sequence extending from nucleotide 10 to nucleotide 28 of the mouse FATP4 coding sequence. FATP4-AS2MM is a control oligo; in the oligo every third nucleotide was changed creating mismatches; the overall nucleotide composition is identical to  
 20 FATP4-AS2 (same number of G, A, T, C). FATP4-S2 is the sense control.

Enterocytes were isolated from the small intestine of mice and incubated for 48h in tissue culture (Figure 40) either without oligonucleotides (squares) or with 100  $\mu$ M FATP4 specific sense (circles) or antisense (diamonds) oligonucleotides. The uptake over time of 25  $\mu$ M oleate was then measured. While the FATP4 sense  
 25 oligonucleotide did not significantly influence the uptake, the antisense oligonucleotide inhibited fatty acid uptake by ~50%.

The effect of either FATP4 sense, antisense or mismatch sequence oligonucleotides on the uptake of fatty acids was measured in enterocytes. Isolated enterocytes were incubated with increasing concentrations of FATP4 antisense

oligonucleotides (solid bars in Figure 41), or a mismatch control oligonucleotide with identical nucleotide composition (stippled bars), or with 100  $\mu$ M of the FATP4 sense-oligonucleotide (lined bar). The medium for this incubation was Dulbecco's modified Eagle's medium with 4.5 g/L glucose, 1 mM sodium pyruvate, 0.01 mg/ml human transferrin and 10% fetal bovine serum. After 48 hours of incubation the uptake of oleate by enterocytes was measured over a 5 minute time interval. Measurements were done in quadruplicate. The uptake assay was done in Hank's buffered salt solution with 10 mM taurocholate. Only the enterocytes given FATP4 antisense oligonucleotide showed a concentration dependent decrease of fatty acid uptake, inhibiting it at a 100  $\mu$ M concentration by ~50%. This effect was FATP4 specific, since only the antisense oligonucleotide which can bind to the FATP4 mRNA and block its translation inhibited uptake, but not a control oligonucleotide differing only in the sequence but not the nucleotide content, ruling out a toxic or otherwise nonspecific inhibitory effect of this oligonucleotide due to its chemical composition.

As a further control experiment, the uptake of oleate was measured along with the uptake of methionine in the same cultured enterocytes. Antisense oligonucleotide, mismatch sequence oligonucleotide, or no oligonucleotide was added to a concentration of 100  $\mu$ M to cultures of enterocytes. After incubation for 48 hours, the uptake of both  $^3$ H-labeled oleate and  $^{35}$ S-labeled methionine was assayed. Results are shown in Figure 42. Fatty acid uptake is at the left side of the paired bars; methionine uptake is on the right side of the paired bars. The fact that amino acid uptake was not influenced by the antisense oligonucleotide treatment further supports the conclusion that the antisense oligonucleotide causes a specific reduction in translation of FATP4-specific mRNA.

#### Example 14: mmFATP2 Is Expressed in Proximal Renal Tubule Epithelium

Northern analysis showed that mmFATP1, mmFATP2, and mmFATP4 are present in the kidney. In situ hybridization (methods as for Example 6) was performed to determine which cell type(s) of the kidney these mRNAs are expressed in. mmFATP1 mRNA was present in virtually all cells throughout the kidney with

no obvious preference for a particular cell type. In contrast, mmFATP2 was expressed only in the renal cortex. Within the cortex, expression of mmFATP2 was restricted to the epithelial cells of the proximal renal tubules. The primary function of proximal renal tubule cells is the reabsorption of filtered salts and nutrients (e.g.,  
5 glucose), a process that requires mitochondrial oxidation and that can utilize fatty acids as energy substrates. Based on the localization of mmFATP2, it is possible that mmFATP2 is important for reabsorption in the kidney by allowing uptake of an energy source (fatty acids) from the blood into renal epithelial cells. Alternatively, if  
10 fatty acids need to be reabsorbed in the kidney, similarly to glucose, FATP2 could be involved in the reabsorption of fatty acids. Determination of the subcellular localization of FATP2 will distinguish between these two possibilities.

Table 5. Mouse FATP mRNA Expression

Mouse Probes	mFATP1	mFATP2	mFATP3	mFATP4	mFATP5
E18.5 embryo expression	everywhere, brain = thymus> heart> brown fat, others	liver (hepatocytes)	-	Brain, small intestine, superior cervical ganglion (SCG), dorsal root ganglion (DRG), other regions have lower expression	Mouse Probes
Duodenum	-	villi (surface epithelium)	-	villi (surface epithelium)	-
Jejunum	-	villi (surface epithelium)	-	villi (surface epithelium)	-
Ileum	-	villi (surface epithelium)	-	villi (surface epithelium)	-
Colon	low expression in the crypt	very low level in the crypt	-	-	-
Kidney	cortex and medulla	proximal tubules	-	-	-

Table 5 (continued). Mouse FATP mRNA Expression

Mouse Probes	mFATP1	mFATP2	mFATP3	mFATP4	mFATP5
Liver	-	hepatocytes	hepatocytes	-	hepatocytes
Pancreas	exocrine secretory units or acinar cells; endocrine pancreas (islet) are negative	exocrine secretory units or acinar cells; endocrine pancreas (islet) are negative	-	-	-
Brain	Neuronal expression throughout the brain including hypothalamus	-	-	Neuronal expression throughout the brain including hypothalamus	-
Heart	myocytes	-	-	-	-
Testis	seminiferous tubules	-	seminiferous tubules	-	-
Lung	bronchiole	-	-	-	-
Adipose	adipocyte	adipocyte	-	-	-

5

## 10 Example 15: Isolation of full-length human FATP3

Full-length clones encoding human FATP3 were identified by searching databases for sequences similar to the murine FATP1-5 coding regions using the BlastX algorithm (Altschul *et al.*, *J. Mol. Biol.* 215: 403-410, 1990). Human clones with similarity to the 5' end of murine FATP sequences were sequenced completely. A clone encoding full-length human FATP3 was



obtained from a human bone library constructed in the mammalian expression vector pMET7 (Tartaglia, L.A. *et al.*, *Cell* 83: 1263-1271, 1995). To identify human cDNA clones encoding FATP family members, databases were searched for sequences similar to murine FATP1-5 coding regions. One clone was found to encode the

5 human ortholog of mmFATP3 and was designated hsFATP3. The DNA and predicted protein sequences of hsFATP3 are shown in Figures 94A and 94B. hsFATP3 is predicted to encode a 702 amino acid 75.6 kD protein with multiple membrane-spanning domains. A comparison of the DNA sequences of mouse and human FATP3 shows that the mouse and human orthologs are 81% identical to each

10 other within the coding region. At the amino acid level, hsFATP3 is ~86% identical to mm FATP3 within the coding region. The sequence identities between mouse and human FATP3 are considerably higher than those observed between different FATP family members within one species (~40%) and are present in the N-terminal part of the protein, a region that is poorly conserved between different FATP family

15 members.

#### Example 16: Substrate Specificity of Fatty Acid Transport in hsFATP-Transfected Clones

Using a mammalian expression vector, we generated 40 stable 239 cell lines expressing hsFATP4 and 20 cell lines transfected with a control plasmid. The ability

20 of the different cell lines to take up FA, as assessed by uptake assays using the fluorescently labeled Bodipy-palmitate, correlated well with their FATP4 expression levels determined by Western blotting (FIG. 95). All 20 vector control clones showed amounts of Bodipy-FA uptake similar to each other and to untransfected 239 cells. In contrast, among the 40 FATP4 transfected clones, a large number (~20) showed an

25 approximately 2-fold increase in Bodipy-FA uptake compared to any of the vector controls, and three had a 5- to 10-fold increase in Bodipy-FA uptake.

Several of the cell lines with the highest amount of Bodipy-FA uptake as well as isolated primary enterocytes were used to measure the uptake of radiolabeled FAs. Short-term uptake by 293 cells and enterocytes of all FAs tested was linear (FIG. 97).

30 hsFATP4 expression enhanced the rate of palmitate uptake approximately 3 fold over

293 cells transfected with vector alone (FIG. 97) and also accelerated the uptake of oleate but not of linolate, arachidonate, octanoate, butyrate or cholesterol (Table 6). Isolated primary enterocytes showed a similar preference for palmitate and oleate, and absence of transport of arachidonate, octanoate, and butyrate, but displayed a more robust transport of linolate and cholesterol than the transfected 293 cells.

To further characterize the substrate specificity of FATP4, we measured the uptake by stably transfected 293 cells of 5  $\mu$ M Bodipy-FA in the presence of a 20 fold molar excess (i.e., 100  $\mu$ M) of FAs, FA-derivatives and lipid soluble vitamins and hormones. Both saturated and non-saturated fatty acids containing 10 to 26 C atoms strongly competed for uptake of Bodipy-palmitate (FIG. 96 and Table 7) and thus are presumed to be substrates of FATP4. In contrast, fatty acids with eight or fewer C atoms did not compete and thus are presumed not to be FATP4 substrates. Similarly, esters of long chain FAs and other hydrophobic molecules tested had no effect on uptake of Bodipy-palmitate.

#### LCFA Uptake Assays (Methods)

Bodipy-FA uptake assays using FACS were performed, adapted to a 96-well format. LCFA uptake assays with enterocytes or with stably transfected 293 cells were done as follows. Mixed micelles of radiolabeled FA (NEN) and taurocholate (Sigma) in HBS were generated by brief sonication at 37°C. Equal volumes of cells and micelle solution were mixed, resulting in a final FA concentration of 25  $\mu$ M for antisense assays and 10  $\mu$ M for substrate specificity assays. Final taurocholate concentration was 5 mM. Cells were incubated for the indicated amount of time at 37°C. The assay was stopped by transferring the cells onto filter paper followed by extensive washes with ice-cold HBS containing 0.1% BSA using a cell harvester (Brandell). Incorporated oleate was then determined by  $\beta$ -scintillation counting (Beckman).

Table 6

Uptake of Different Substrates by FATP4 Expressing Cell Lines and  
Enterocytes

Fatty Acid	293 Cells Control*	293 Cells Stably Expressing FATP4	FATP4 specific	Enterocytes*
Palmitate	564	1695	1131	3036
Oleate	662	1122	459	117
Linolate	640	673	33	116
Arachidonate	3	5	2	0
Octanoate	0	0	0	5
Butyrate	0	50	50	73
Cholesterol	319	345	26	531

Uptake of different substrates by enterocytes and by control and stable FATP4-expressing 293 cells. The rates of uptake for the indicated fatty acids was measured over 4 min taking measurements every 30 s. All fatty acids were at a concentration of 10  $\mu$ M in HBS containing 5 mM taurocholate.

\*Uptake measured as *pmol/min 10<sup>6</sup> cells*

Table 7

Competition of Bodipy-FA Uptake by FATP4 Expressing Cells

Fatty Acids	Formula	Competition
Butyric Acid	C <sub>4</sub> H <sub>8</sub> O <sub>2</sub>	-

-93-

5	Caproic Acid	$C_6H_{12}O_2$	-
	Caprylic Acid	$C_8H_{16}O_2$	-
	Capric Acid	$C_{10}H_{20}O_2$	++
	Lauric Acid	$C_{12}H_{24}O_2$	++
	Myristic Acid	$C_{14}H_{28}O_2$	++
10	Palmitic Acid	$C_{16}H_{32}O_2$	++
	Stearic Acid	$C_{18}H_{36}O_2$	+
	Oleic Acid	$C_{18}H_{34}O_2$	++
	Linoleic Acid	$C_{18}H_{32}O_2$	++
	Arachidic Acid	$C_{20}H_{40}O_2$	++
	Lignoceric Acid	$C_{24}H_{48}O_2$	++
	Cerotic Acid	$C_{26}H_{52}O_2$	++

## Fatty Acid Derivatives

15	Fatty Acids	Formula	Competition
	Palmitic Acid Methyl Ester	$C_{17}H_{34}O_2$	-
	Stearic Acid Methyl Ester	$C_{19}H_{38}O_2$	-
	Oleic Acid Ethyl Ester	$C_{20}H_{38}O_2$	-
	Oleic Acid Oley Ester	$C_{36}H_{68}O_2$	-
20	Oleoyl CoA	$C_{39}H_{68}N_7O_{17}P_3S$	-
	Cholesteryl Oleate	$C_{45}H_{78}O_2$	-

Table 7 Continued

Competition of Bodipy-FA Uptake by FATP4 Expressing Cells

Lipid-Soluble Vitamins &amp; Hormones

	Fatty Acids	Formula	Competition
	Retinoic Acid (Pro-Vitamin A)	$C_{20}H_{28}O_2$	$\pm$
	Ergocalciferol (Vitamin D2)	$C_{28}H_{44}O_2$	-
	Tocopherol (Vitamin E)	$C_{29}H_{50}O_2$	-
5	3-Phytylamenadione (Vitamin K1)	$C_{31}H_{46}O_2$	-
	Prostaglandin E2	$C_{20}H_{32}O_5$	-

Competition for Bodipy-FA uptake by FATP4 expressing cells by different hydrophobic compounds. The uptake of 5  $\mu$ M Bodipy-FA, C1-Bodipy-C12 was measured in the presence of a 20-fold molar excess (i.e., 100  $\mu$ M) of the indicated fatty acids or fatty acid derivatives. The maximal 100% inhibition was defined as the amount of Bodipy-FA incorporated in the presence of 200  $\mu$ M lauric acid which was on average  $18\% \pm 5\%$  that of untreated cells.

- 15    -: 0% - 30% inhibition by the indicated substance  
       $\pm$ : 30% - 50% inhibition  
      +: 50% - 70% inhibition  
      ++: 70% - 100% inhibition

## Example 17: Identification and Characterization of the FATP5 Promoter

## METHODS

## BAC Isolation and Luciferase Constructs

An arrayed BAC library was screened by PCR for FATP5 genomic clones.

- 5 PCR primers designed by a program from the Whitehead Institute's Genome Center specifically amplified a single band of the correct size from mouse genomic DNA. Two putative BACs containing the FATP5 genomic sequence were identified and the presence of FATP5 sequence was confirmed by dot hybridization of the BAC with the mmFATP5 cDNA.

- 10 After isolation of positive BACs, large amounts of bacteria were grown and DNA prepared using a Qiagen maxi-prep kit (Qiagen, Venlo, The Netherlands). The BAC was digested with Sac I and ligated into pZero-2 (Invitrogen, Carlsbad, CA). Inserts containing mmFATP5 genomic sequence were identified by screening colony lifts of the ligation with an  $\alpha$ -<sup>32</sup>P-ATP radiolabeled, random primed (Boehringer-
- 15 Mannheim, Indianapolis, IN) mmFATP5 cDNA as a probe. Positive colonies were picked and restriction analysis with Sac I revealed them to contain an identical, large insert of 8-10 kb. Digestion of the Sac I fragment with BstX I yielded three pieces that were subsequently subcloned into pZero and sequenced using an ABI sequencer (Research Genetics). A 1.3 kb piece containing sequence immediately upstream of
- 20 the FATP5 initiator methionine was subcloned into the Xho I and Bgl II sites of the promoter-less pGL3 luciferase reporter vector (Promega Corp., Madison, WI). 7 kb of additional upstream sequence was subcloned into the Xho I and Sac I sites of the prior construct to yield a final construct containing approximately 8 kb of genomic sequence upstream of the initiator methionine. Deletions of the FATP5 promoter
- 25 were constructed using PCR with the 1.3 promoter construct as the template. Products were amplified with primers containing Hind III (5' primer) and Xho I (3' primer) sites using Elongase (Gibco, Rockville, MD). The resulting fragments were cut with Hind III and Xho I and subcloned into the corresponding sites of the promoter-less pGL3 luciferase reporter vector. The internal 30 base pair deletions,
- 30 GC box mutations, and 10 nucleotide linker scan were all created with the

Quickchange mutagenesis kit (Stratagene, La Jolla, CA) according to the manufacturer's instructions. At least two different bacterial colonies were picked for each construct. The inserts from both colonies were sequenced to check for unintended point mutations and both constructs were assayed for luciferase activity.

## 5 Cell culture, Transfection, and Luciferase Measurements

HepG2, Hep3B, HT1080, 3T3-L1, BOSC, and HACAT cells were grown in DMEM supplemented with 10% fetal calf serum, 1 x penicillin-streptomycin and glutamine (Gibco, Rockville, MD). Mink lung cells were grown in MEM supplemented with 10% fetal calf serum, 1 x minimal essential amino acids, 1 x penicillin-streptomycin and glutamine. The evening prior to transfection, cells were  
10 plated at 50-60% confluence in 24 well dishes. The following morning, cells were placed in 2 mls of fresh media and 250  $\mu$ L of a  $\text{CaPO}_4$  solution (Invitrogen, Carlsbad, CA) containing 2  $\mu$ g of a luciferase reporter construct and 0.5  $\mu$ g of pCMV- $\beta$ -gal was added to the cells. pCMV- $\beta$ -gal constitutively expresses  $\beta$ -galactosidase and  
15 was used to normalize transfection efficiency (Hua et al., 1998). After 12 hours, the cells were washed twice with DMEM and placed in fresh media. Thirty six hours later, the media over the cells was removed and 250  $\mu$ L of 1 x reporter lysis buffer (Promega Corp., Madison, WI) was added. After vigorous shaking for 15 minutes at room temperature, the supernatants were transferred to Eppendorf tubes and briefly  
20 centrifuged to remove particulates. 20  $\mu$ L from these tubes was used for determination of luciferase activity (Promega Corp., Madison, WI) and 20  $\mu$ L was used for the measurement of  $\beta$ -galactosidase activity (Clontech, Palo Alto, CA). All luciferase values were normalized to  $\beta$ -galactosidase to control for transfection efficiency and expressed as relative luciferase units (RLU). For experiments  
25 comparing different cell lines, promoter activity was computed as a fold induction by dividing the RLU activity of either the -8 or -271 promoter constructs by the RLU activity a promoter-less construct. Each data point was done in triplicate and each experiment was repeated a minimum of three times.

-97-

### Northern Blots, Preparation of Nuclear Extracts, and Gel Shift Assays

Human poly-A northern blots were purchased from a commercial vendor (Clontech, Palo Alto, CA) and probed with a piece of the human FATP5 3' untranslated region specific for FATP5. Nuclear lysates from HepG2 and BOSC  
5 cells were essentially prepared according to the method of Hua et al. and stored at -80°C (Hua et al., 1998). Probes for gel shift assays were end labeled using T4 polynucleotide kinase (Boehringer-Mannheim, Indianapolis, IN) and gel purified. Gel shifts were performed at room temperature in 30 µL reactions comprised of 6 µL  
10 5 X binding buffer (100 mM Tris 8.0, 300 mM KCl, 5 mM EDTA, 8 mM MgCl<sub>2</sub>, and 36% glycerol), 0.5 µL of 100 mM DTT, 1 µL of 10 mg/ml BSA, 2 µL of 2 mg/ml poly dI/dC, and 5 µL nuclear lysate. Ten minutes after the addition of nuclear lysate, 40,000 cpm of <sup>32</sup>P-labeled probe were added. After 20 minutes at room temperature, loading dye was added and the reaction run on a 4% non-denaturing gel.

### RESULTS

15 Human FATP5 mRNA is only expressed in adult liver

We had previously reported that mmFATP5 mRNA was only expressed in the liver (Hirsch et al., 1998). To determine if the human isoform of FATP5 was also liver specific, we performed northern analysis using a probe from the 3' transcribed but untranslated region of the human gene. Similar to the mouse homolog, hsFATP5  
20 is liver specific. Interestingly, hsFATP5 was not expressed in fetal liver suggesting that it may be developmentally regulated.

### Identification of a FATP5 promoter

We next set out to determine the cis-acting elements responsible for liver specific expression of FATP5. We identified BACs containing the FATP5 genomic  
25 locus and subcloned a 10 kb Sac I fragment which was subsequently sequenced. The Sac I fragment contains approximately 8 kb of genomic sequence upstream of the FATP5 initiator methionine. Blast searches using the 5' end of the Sac I sequence



-98-

revealed that it contained coding sequence for an unknown gene immediately upstream of FATP5. Since the FATP5 promoter is unlikely to overlap the coding sequence of another gene, we hypothesized that the 10 kb Sac I fragment contained the FATP5 promoter. To test this hypothesis, 8 kb of genomic DNA upstream of the translational initiator of FATP was subcloned into the promoter-less pGL3 luciferase reporter vector. This construct was transiently transfected into the HepG2 liver cell line and luciferase activity was determined. The -8 kb piece of DNA resulted in a 35 fold induction of luciferase activity when compared to a pGL3 vector without the FATP5 genomic sequence (FIG. 100). To determine if this activity reflected tissue specific transcription, the -8 kb luciferase reporter construct was transfected into a variety of additional cell types. While promoter activity was also detected in the Hep3b hepatoma cell line, non-liver cell lines did not express luciferase above the level of the promoter-less vector. Thus, the 8 kb upstream genomic element recapitulated liver specific expression *in vitro*.

15 The FATP5 promoter resides within the 261 base pairs upstream of the initiator methionine and requires a single GC box

To determine the cis-acting elements in the -8 kb of genomic sequence responsible for transcriptional activity, serial 5' deletions of the promoter were constructed and transfected into HepG2 cells. Surprisingly, greater than 90% of the -8 kb was dispensable for promoter activity. A construct containing only 261 base pairs upstream of the initiator methionine resulted in promoter activity equivalent to that of the -8 kb construct (FIG. 101). Identical results were obtained when the deletion series was transfected into Hep3b cells (data not shown). We next determined if promoter activity of a small genetic element was tissue specific. Transfection of a construct containing 271 base pairs upstream of the initiator methionine into a variety of cell lines essentially replicated the results of the -8 kb construct in that expression was observed only in liver derived cell lines (FIG. 102).

Since deletion analysis revealed that bases between -261 and -218 were required for promoter activity, we closely examined this region for binding sites of known transcription factors and found the sequence GGGGCGGGG between

nucleotides -241 and -232 (FIG. 103A). This sequence binds the Sp1 family of transcription factors and is termed a GC box. To determine if the activity of the -271 construct required the GC box, we mutated the GC box. The first construct deleted nucleotides -241 to -222 which removed the GC box and additional downstream  
5 sequence which, although less optimal, might also bind the Sp1 family of transcription factors (SEQ ID NO.: 107). The second construct had three G to A point mutations in the GC box between nucleotides -241 to -232 (SEQ ID NO.: 108). Such mutations had previously been shown to abolish transcriptional activity of GC boxes (Rodenburg et al., 1997). In contrast to the wild type -271 promoter, both of the  
10 mutated constructs were transcriptionally inactive in HepG2 cells (FIG. 103B). Identical results were also obtained in Hep3B cells (data not shown). This suggests that the GC box between -241 to -232 is essential for transcriptional activity of the FATP5 promoter. We next examined whether the sequences necessary for luciferase activity also bound proteins in nuclear extracts from HepG2 cells. Two different  
15 oligonucleotides were used for gel shift analysis. One oligonucleotide (AF-1) contained nucleotides -250 to -230 (SEQ ID NO.: 111) and the other (AF-2) spanned nucleotides -260 to ~-200 (SEQ ID NO.: 109) (FIG. 104). Both oligonucleotides yielded three significant complexes from HepG2 nuclear extracts. All complexes were specific as 100 fold excess of the same unlabeled oligonucleotide could compete  
20 for binding of the radiolabeled oligonucleotide. Mutant AF-1 oligonucleotides containing three point mutations in the GC box did not bind any proteins in HepG2 nuclear extracts or compete for binding of nuclear proteins to the AF-1 or AF-2 oligonucleotides (data not shown). Oligonucleotides AF-1 and AF-2 also bound recombinant Sp1 (Promega Corp, Madison, WI, data not shown). However, nuclear  
25 extract from BOSC cells, a kidney cell line, and HepG2 cells had identical patterns of complex formation (data not shown).

#### Identification of novel sequences required for transcriptional activity of the FATP5 promoter

While the GC box between nucleotides 241 and 232 is essential for  
30 transcriptional activity, additional sequences downstream of the GC box might also

-100-

be required for transcription. To determine if such sequences existed, we created 30 base pair internal deletions in the ~-271 construct downstream of the GC box. Constructs that had deletions in sequences between 240 and 180 nucleotides upstream of the FATP5 translational initiator had greatly reduced transcriptional activity in

5 HepG2 cells (FIG. 105). To identify the specific sequences within this region required for FATP5 transcription, a 10 nucleotide linker (CTAACAGGAG) (SEQ ID NO.: 113) was exchanged for wild type sequence within the context of the -271 base pair construct (FIG. 106). Inadvertently, the 210 to 200 construct had a single nucleotide insertion and the 190 to 180 construct had a two nucleotide insertion

10 relative to the wild type sequence. However, several other linker constructs that also had equivalent insertions (230 to 220 or 170 to 160 for example) had high levels of luciferase activity. Thus the decrease in luciferase activity in the 190 to 180 and 210 to 200 constructs is due to changes in the nucleotide sequence and not the result of the nucleotide additions. Transfection of these DNA into HepG2 cells revealed two

15 regions important for transcription. Mutating sequences between nucleotides -210 and ~-200 or between nucleotides -190 and -180 drastically reduced luciferase activity (FIG. 106).

In both humans and mice, FATP5 is only expressed in the liver. To determine the promoter elements mediating liver specific transcription, we isolated a BAC

20 encoding the mouse FATP5 genomic locus and sequenced 10 kb upstream of the transcriptional start. Since this 10 kb of genomic DNA did not contain either a TATA box or GC rich regions found in TATA-less promoters, FATP5 may utilize non-canonical sequences for transcription initiation. Unfortunately, attempts to identify the transcriptional start using primer extension were unsuccessful, perhaps

25 due to secondary structure in the 5' UTR. Since we did not unambiguously determine the transcriptional start site, the nucleotide numbering in all of the promoter constructs refers to the distance from the translational start codon.

#### GC box and Sp1 transcription factors

Since another gene was situated approximately 8 kb upstream of the FATP5

30 initiator methionine, we hypothesized that promoter elements were likely within this

region of DNA. A luciferase reporter construct containing this sequence was transcriptionally active in two liver cell lines but was inactive in cell lines derived from lung, muscle, kidney, skin, or fibroblasts. Deletion analysis of the -8 kb reporter construct revealed that the FATP5 promoter was contained within the 261 nucleotides upstream of the initiator methionine. Promoter activity in this -261 base pair piece required the presence of a single GC box. Gel shift assays with oligonucleotides containing this GC box revealed the presence of three distinct complexes that required a functional GC box for binding. GC boxes bind the Sp1 family of transcription factors and the multiple complexes could reflect the binding of different members of the Sp1 protein family or different post-translational modifications of Sp1 in HepG2 cells (Rodenburg et al., 1997). Although the Sp1 family of transcription factors is widely expressed, Sp1 has been shown to be important for the transcription of several liver specific genes and is upregulated in liver after birth (Rodenburg et al., 1997). In some cases, Sp1 will facilitate the binding of a tissue specific transcription factor to DNA. For example, Sp1 binding to DNA enhances the binding of C/EBP $\beta$  to an adjacent site in the liver specific CYP2D5 promoter (Lee et al., 1994). Since the C/EBP $\beta$  binding site in the CYP2D5 promoter is suboptimal, C/EBP $\beta$  binding to this site requires the presence of Sp1 or nuclear extract. A similar situation could occur in the FATP5 promoter. Although mutations in the 10 nucleotides downstream of the GC box had no effect on luciferase activity, we did not test mutations immediately upstream of the GC box for effects on promoter activity. It is also possible that Sp1 might bind an unknown liver specific transcription factor and recruit it to the FATP5 promoter. Although, there is no experimental evidence for this, Sp1 has recently been shown to bind to a transcriptional activator so additional interacting proteins are possible (Ryu et al., 1999).

#### Other liver specific transcription factors

Alternatively, since the Sp1 gene family is important for the transcription of many genes which are not liver specific, liver specific promoter elements in the FATP5 promoter might be located elsewhere (Boisclair et al., 1993; Rongnoparut et al., 1991; Sorensen and Wintersberger, 1999). Analysis of the sequence downstream

of the GC box using TFSearch  
(<http://pdap1.trc.rwcp.or.jp/research/db/TFSEARCH.html>) did not reveal any  
additional transcription factor binding sites of relevance (Heinemeyer et al., 1999;  
Heinemeyer et al., 1998). Further, we were unable to visually identify binding sites  
5 for known liver specific transcription factors in this sequence (De Simone and  
Cortese, 1992; Hanson and Reshef, 1997; Lai, 1992). Thus, we looked  
experimentally for additional promoter elements by mutating the sequence  
downstream of the GC box and identified two additional sites downstream of the GC  
box that were essential for FATP5 transcription. The sequences of these sites do not  
10 conform to any known transcription factor binding sites suggesting the either novel  
proteins bind these elements or that these elements bind known proteins in a novel  
manner. Preliminary gel shift data using oligonucleotides spanning these site  
suggests that these two elements may comprise a binding site for a single complex.  
Further additional data suggests that the complex which binds to these two sites  
15 interacts with the GC box 30 base pairs upstream. Interestingly, we noted a  
palindromic sequence equally split between these two sites (FIG. 107). Since many  
transcription factors bind palindromic DNA elements, it is intriguing to speculate that  
these two sequences contribute to the binding site for a novel transcription factor.  
Current investigations are focused on identifying the proteins binding to these novel  
20 elements and how this element interacts with the GC box.

— Several studies have shown that the FATP gene family is regulated by a  
variety of substances including LPS, cytokines, insulin, and diet (Frohnert et al.,  
1999; Hui et al., 1998; Memon et al., 1999). Especially intriguing has been a recent  
report that FATP1 is upregulated by PPAR $\alpha$  ligands in liver cell lines (Martin et al.,  
25 1997; Motojima et al., 1998). Since fatty acids may be endogenous activators of  
PPAR's, transcriptional regulation of FATP1 by PPAR's may represent a physiologic  
feedback loop (Gottlicher et al., 1992; Grimaldi et al., 1999; Schoonjans et al., 1996).  
Given that liver also expresses FATP5, it will be interesting to see whether this genes  
is also regulated by PPAR $\alpha$  and the tools developed here should help address this  
30 question.

-103-

Several factors make the FATP5 promoter amenable to further study. First, liver specific transcription of FATP5 can be recapitulated using immortalized cell lines *in vitro*. Second, the minimal required promoter element that confers liver specific transcription is very small. Third, transcriptional activity of this promoter is very robust. Thus, further study of the FATP5 promoter may provide additional insight into the mechanisms of liver specific transcription and regulation of the FATP gene family.

#### Example 18:

##### Materials and Methods

Polyclonal antibodies were raised against proteins containing the N-terminal domain of mouse FATP2 or the C-terminal domain of mouse FATP5 fused to glutathione-S-transferase (GS). Tissues for immunofluorescence were collected from 8 week old mice and a 2 year old chimpanzee. Tissues were fresh frozen, cut on a cryostat and mounted on slides. Immunofluorescence was performed as previously described (Stahl et al., 1999). Pictures were taken on a Zeiss confocal microscope.

To determine FATP2 expression in the gall bladder, mouse gall bladder was incubated with anti-FATP2 antibody as the primary antibody and rhodamine-labeled anti-rabbit IG as the secondary antibody. FATP2 antibody clearly stained the gall bladder epithelium, but did not result in significant staining of other cell types. (Figure 108)

To further study FATP2 expression, chimpanzee liver was costained with anti-FATP2 antibody(green) and anti CD31 antibody(red). CD31 is expressed on endothelial cells and is used as a marker for blood vessels. FATP2 immunoreactivity was present in large patches which overlap with CD31 positive areas, suggesting that FATP2 protein was present in the space of Diss, the area where hepatocytes exchange nutrients with the blood. This implicates FATP2 in the uptake of fatty acids into hepatocytes. In addition to areas which overlap with CD31 immunoreactivity, FATP2 protein was also present on the cell surface of hepatocytes in a small bead

pattern. Immunoelectronmicroscopy of similar sections showed that FATP2 immunoreactivity was localized in the walls of bile caniculi which are formed by the liver cells. (Figure 109) The presence of FATP2 in bile caniculi in the liver as well as its presence in the gall bladder epithelium suggests a role for FATP2 in either  
5 absorption or secretion of fatty acids into the bile. The levels of free fatty acids in the bile have been associated with the frequency of all stone formation.

To further study FATP5 expression, chimpanzee liver was costained with anti-FATP5 antibody(green) and anti CD31 antibody(red). CD31 is expressed on endothelial cells and is used as a marker for blood vessels. FATP5 immunoreactivity  
10 was present in large patches which overlap with CD31 positive areas, suggesting that FATP5 protein was present in the space of Diss, the area where hepatocytes exchange nutrients with the blood. (Figure 110) This implicates FATP5 in the uptake of fatty acids into hepatocytes.

#### Example 19 Identification and Characterization of Human FATP3 Proteins

##### 15 Isolation of additional humanFATP3 clones

An additional clone encoding human FATP3 was identified by searching for sequences similar to murine or human FATP3 coding regions using the BlastX algorithm in a proprietary database, (Altschul, et al, J. Mol. Bio. 215: 403-410, 1990). One clone, which was identified by random library sequencing, is described as  
20 johni003f04 (SEQ ID NO: 116) extends the open reading frame of the hsFATP3 polypeptide sequence by 30 amino acids at the N-terminus when compared to previously discovered sequences. The DNA sequence of this clone is shown in Figures 111A and 111B, and the predicted protein sequence (SEQ ID NO: 117) is shown in Figure 112. The open reading frame of this clone begins at the initial  
25 nucleotide and includes nucleotide 2240. The first ATG is located at nucleotide number 51, resulting in a predicted protein which includes 730 amino acids. An FATP signature sequence (see Hirsch et al., PNAS, 95:8625-8629, 1998) is clearly

present between amino acids 331 and 640 of hsFATP3. Within this signature sequence hsFATP3 is 48% identical to hsFATP1 at the amino acid level. A consensus AMP-binding motif has been identified (amino acid 333-334). Thus, hsFATP3 is clearly a member of the fatty acid family.

## 5 Functional analysis of FATP3 Clones

SEQ ID NO: 116 is contained in the mammalian expression vector pMET7 (Tartaglia, *et al.*, Cell, 83: 1263-1271, 1995). To determine if the protein encoded by this DNA sequence can mediate fatty acid uptake, SEQ ID NO: 116 was transfected into COS cells. Uptake of a BODIPY-labeled fatty acid was determined as described  
 10 in previous experiments (Hirsch, *et al.*, PNAS, 95: 8625-8629, 1998). Transfection with SEQ ID NO: 116 resulted in a dramatic increase in fatty acid uptake when compared to transfection with vector control. In this experiment, CD31 served as a marker for transfected cells. Only CD31 positive cells were considered for analysis (see Hirsch, *et al.*, PNAS, 95: 8625-8629, 1998 for details). The results (Figure 113)  
 15 demonstrate that SEQ ID NO: 116 encodes a functional fatty acid transport protein.

## Tissue Distribution of human FATP3

Polyclonal antibodies were raised by immunizing rabbits with GST fused to the most C-terminal 89 amino acids of mmFATP3 -  
 (RPPQALNLVQLYSHVSENLPYARPRFLRLQESLATTETFKQKQVRMANEGF  
 20 DPSVLSDPLYVLDQDIGAYLPLTPARYSALLSGDLRI) (SEQ ID NO: 120). Western blotting experiments with murine tissue lysates using the anti-FATP3 antiserum closely confirmed the unique expression pattern of FATP3 as judged by northern blot experiments. This, together with the fact that the serum reacted only weakly with lysates from cell lines expressing either FATP1, -2, -4 or -5, indicates  
 25 that the antibody recognizes preferentially FATP3, but not other FATP family members.

FATP3 protein was detected in mouse liver, spleen, heart, kidney, testis, white adipose tissue, and most notably in the lung. Further FATP3 expression in the lung was examined by immunofluorescence microscopy. 5 to 10  $\mu$ M thick fresh frozen



unfixed sections of murine and chimpanzee lungs were blocked with 10% FCS/1% donkey serum/1% BSA in HBS and incubated overnight with anti-FATP3 serum in blocking solution. After washing the sections Alexa 488 conjugated donkey anti-rabbit secondary antibodies were used to detect bound anti-FATP3 primary antibodies and nuclei were stained TOTO3. In later experiments, chimpanzee lung was incubated with a mixture of rabbit anti-FATP3 and mouse monoclonal anti-CD31 to visualize FATP3 as well as blood vessels. Sections were imaged on a Zeiss LSM510 confocal microscope. Experiments carried out once with mouse and three times with chimpanzee lung tissue showed that FATP3 is present at high levels in type-II pneumocytes, a cell type responsible for secretion of surfactant, a phospholipid-rich film critical for lung function. The exact function of FAT3 in type II pneumocytes is not yet clear. One hypothesis is that FATP3 is responsible for supplying fatty acid substrates for the synthesis of surfactant.

PCR-based experiments showed that the exocrine as well as endocrine pancreas expresses FATP3. This fact was confirmed by immunofluorescence performed as described above for the lung sections, on chimpanzee pancreas which showed FATP3 localized to the plasma membrane of acinar cells and a punctate expression pattern on the plasma membrane and in the cytosol of alpha and beta cells of the pancreatic islands. The identification of a fatty acid transporter in the insulin producing cells of the pancreas has potentially broad implications for the treatment of type II diabetes and obesity. In both diseases, fatty acid levels in the blood are elevated and, in later stages of the disease, lead to diminished insulin secretion by the pancreas due to the induction of apoptosis in insulin-producing beta cells (Shimabukuro, *et al.*, PNAS, 95: 2498-2502, 1988). Blocking fatty acid uptake into the beta cells could possibly prevent apoptosis and maintain insulin secretion thus preventing the progression from obesity to diabetes.

#### Example 20 Identification of a fatty acid binding domain in FATP4

GST fusion proteins were constructed in pGEX for four regions of hsFATP4 (SEQ ID NO: 52; Figure 51) which were generated by PCR and verified by sequencing. The first three fusion proteins were constructed from regions near the N-

terminal portion of the protein. SP1 (SEQ ID NO:121) contained amino acid residues 43-239 of the hsFATP4 sequence as shown in Figure 114A. This portion of hsFATP4 contains a lipocalin domain (as shown in Figure 117) as well as a number of residues which in hsFATP4 are upstream of the lipocalin domain. SP2 (SEQ ID NO: 122) contained residues 43-290 of the hsFATP4 sequence as shown in Figure 114B. This portion of the hsFATP4 contains a lipocalin domain and an AMP binding domain as well as a number of residues which are upstream of the lipocalin domain. SP3 (SEQ ID NO: 123) contained amino acid residues 125-290 of the hsFATP4 sequence as shown in Figure 114C). This portion of the hsFATP4 contains a lipocalin domain and an AMP binding domain, but does not contain the upstream residues. The fourth fusion protein was constructed from a region at the C-terminal end of the hsFATP4 polypeptide. SP5 contained amino acid residues 417-643 of hsFATP4 polypeptide as shown in Figure 114D (SEQ ID NO: 124).

Proteins were expressed in *E. coli* and purified on glutathione affinity beads using standard techniques. To determine fatty acid binding, beads were mixed with 100  $\mu$ M <sup>14</sup>C-labeled fatty acids in mixed micelles with taurocholate (10mM, Sigma) and incubated for 30 minutes at room temperature. The beads were subsequently washed with PBS containing 10mM taurocholate and radioactivity associated with beads was assessed by scintillation counting. A fusion to the C-terminal domain of hsFATP4 (SP5) did not show any oleate (ARC) binding compared to GST protein alone, while 2 N-terminal fusions (SP1 and 2) bound significant amounts of oleate. (Figure 116).

FATTY ACID	SP1	SP2	SP3	SP5	GST
Oleate	25772±1326	16172±1639	4206±631	2413±186	1511±525

Similar results were obtained using maltose-binding protein fusions. MBP fusion constructs were generated by digesting the pGEX-SP constructs with EcoRI/XhoI and ligated into pMAL digested with EcoRI/SaII. MBP fusion proteins were expressed in *E. coli* and were purified under non-denaturing conditions following

the manufacturer's instructions. To determine fatty acid binding, beads were mixed with 100  $\mu$ M  $^{14}$ C-labeled fatty acids in mixed micelles with taurocholate (10mM, Sigma) and incubated for 30 minutes at room temperature. The beads were subsequently washed with HBS containing 10mM taurocholate. The proteins were subsequently eluted from the resin with maltose and the amount of fatty acid binding to MBP-SP1, -2, -3, and -5 was assessed by determining the radioactivity associated with the elute by  $\beta$ -scintillation counting.

Unlike GST fusion proteins, MBP fusion proteins are not self-dimerizing. Further, long-chain fatty acids (such as oleate and palmitate), but not short-chain fatty acids (such as butyrate), were specifically bound by SP1 (Figure 117). This selective binding is consistent with previous reports of the substrate specificity of FATP4 (Stahl, *et al.*, Mol. Cell, 4, 299-308, 1999). The identification of a fatty acid binding domain in FATP4 will be useful in the development of small molecules that inhibit the binding and transport of fatty acids by FATP4 and may provide useful information on the mechanism of fatty acid transport.

#### Results of Fatty Acid Binding

FATTY ACID	Composition	binding to MBP-SP1	binding to MBP-SP5
Oleate	C <sub>18</sub> H <sub>34</sub> O <sub>2</sub>	3968	2800
Palmitate	C <sub>16</sub> H <sub>32</sub> O <sub>2</sub>	4588	844
Arachidonate	C <sub>20</sub> H <sub>40</sub> O <sub>2</sub>	1942	1147
Butyrate	C <sub>4</sub> H <sub>8</sub> O <sub>2</sub>	142	633

These experiments demonstrate that the FATPs of the present invention contain domains that bind various long chain fatty acids. Thus, polypeptides containing these domains can be prepared and utilized to assess the modulation of binding and transport function by a variety of agents. The polypeptides with the highest binding capacities were shown to be those containing a lipocalin domain (such as those shown in Figure 118) with additional upstream residues, such as those

associated with this domain in the N-terminal portion of hsFATP4. Polypeptides containing domains in addition to the lipocalin domain (for example, those containing an AMP binding domain) were also shown to bind fatty acids at significant levels.

Figure 118 contains an alignment depicting the consensus sequences for the six  
 5 human FATP, hsFATP1, hsFATP2, hsFATP3, hsFATP4, hsFATP5 and hsFATP6 polypeptides. A lipocalin domain and an AMP binding domain for each polypeptide are both identified and compared. A search using the lipocalin signature sequence [DENG]-X-[DENQGSTARK]-X(0,2)-[DENQARK]-[LIVFY]-{CP}-G-{C}-W-[FYWLRH-X]-[LIVMTA] conducted on a public database ([www.ebi.ac.uk/interpro/](http://www.ebi.ac.uk/interpro/)),  
 10 indicated that the lipocalin domains of hsFATP1 and hsFATP4 are identical to the lipocalin signature sequence. In addition, a search directed to identifying sequences having at least 80% identity to the lipocalin signature sequence identified three additional human FATPs, hsFATP3, hsFATP5 and hsFATP6.

15 The following is the result of comparing individual hsFATP protein sequences with the lipocalin domain identified for hsFATP1 and hsFATP4. The comparison was made using the BLAST Network Service at the National Center for Biotechnology Information. (Capitalized AA agree with the lipocalin signature sequence.)

FATP6: 114 to 125 NEpDFVhVWFGL. 76% similarity (SEQ ID NO: 138)  
 20 AATGAGCCGGACTTCGTTACGTGTGGTTCGGCCTC

FATP5: 182 to 194 sQAVpaLcMWLGL. 53% similarity (SEQ ID NO: 139)  
 TCCCAGGCCGTTCCAGCCCTGTGTATGTGGCTGGGGCTG

FATP4: 134 to 146 ENRNEFVGLWLGM. Identity (SEQ ID NO: 129)  
 GAGAACCGCAATGAGTTCGTGGGCCTATGGCTGGGCATG

-110-

FATP3: 221 to 234 IPAGPEFLwLWFGL. 69% similarity (SEQ ID NO: 140)  
CTCCCCGCTGGCCCAGAGTTTCTGTGGCTCTGGTTCGGGCTG

FATP2: 112 to 124 GNEPAYVwLWLGL. 80% similarity (SEQ ID NO: 127)  
GGTAACGAGCCGGCCTACGTGTGGCTGTGGCTGGGGCTG

5 FATP1: 136 to 148 EGRPEFVGLWLGL. Identity (SEQ ID NO: 126)  
GAGGGCCGGCCGGAGTTCGTGGGGCTGTGGCTGGGCCTG

#### REFERENCES

Abumrad, N., Coburn, C., and Ibrahimi, A. (1999). Membrane proteins implicated in long-chain fatty acid uptake by mammalian cells: CD36, FATP and FABPm. *Biochim*  
10 *Biophys Acta* 1441, 4-13.

Berk, P. D., Bradbury, M., Zhou, S. L., Stump, D., and Han, N. I. (1996). Characterization of membrane transport processes: lessons from the study of BSP, bilirubin, and fatty acid uptake. *Semin Liver Dis* 16, 107-20.

Berk, P. D., and Stump, D. D. (1999). Mechanisms of cellular uptake of long chain  
15 free fatty acids. *Mol Cell Biochem* 192, 17-31.

Boisclair, Y. R., Brown, A. L., Casola, S., and Rechler, M. M. (1993). Three clustered Sp1 sites are required for efficient transcription of the TATA-less promoter of the gene for insulin-like growth factor-binding protein-2 from the rat. *J Biol Chem* 268, 24892-901.

- De Simone, V., and Cortese, R. (1992). Transcription factors and liver-specific genes. *Biochim Biophys Acta* 1132, 119-26.
- Fitscher, B. A., Elsing, C., Riedel, H. D., Gorski, J., and Stremmel, W. (1996). Protein-mediated facilitated uptake processes for fatty acids, bilirubin, and other  
5 amphipathic compounds [see comments]. *Proc Soc Exp Biol Med* 212, 15-23.
- Frohnert, B. I., Hui, T. Y., and Bernlohr, D. A. (1999). Identification of a functional peroxisome proliferator-responsive element in the murine fatty acid transport protein gene. *J Biol Chem* 274, 3970-7.
- Glatz, J. F., Luiken, J. J., van Nieuwenhoven, F. A., and Van der Vusse, G. J. (1997).  
10 Molecular mechanism of cellular uptake and intracellular translocation of fatty acids. *Prostaglandins Leukot Essent Fatty Acids* 57, 3-9.
- Gottlicher, M., Widmark, E., Li, Q., and Gustafsson, J. A. (1992). Fatty acids activate a chimera of the clofibric acid-activated receptor and the glucocorticoid receptor. *Proc Natl Acad Sci U S A* 89, 4653-7.
- 15 Grimaldi, P. A., Teboul, L., Gaillard, D., Armengod, A. V., and Amri, E. Z. (1999). Long chain fatty acids as modulators of gene transcription in preadipose cells. *Mol Cell Biochem* 192, 63-8.
- Hamilton, J. A. (1998). Fatty acid transport: difficult or easy? *J Lipid Res* 39, 467-81.
- Hanson, R. W., and Reshef, L. (1997). Regulation of phosphoenolpyruvate  
20 carboxykinase (GTP) gene expression. *Annu Rev Biochem* 66, 581-611.

- Heinemeyer, T., Chen, X., Karas, H., Kel, A. E., Kel, O. V., Liebich, I., Meinhardt, T., Reuter, I., Schacherer, F., and Wingender, E. (1999). Expanding the TRANSFAC database towards an expert system of regulatory molecular mechanisms. *Nucleic Acids Res* 27, 318-22.
- 5    Heinemeyer, T., Wingender, E., Reuter, I., Hermjakob, H., Kel, A. E., Kel, O. V., Ignatieva, E. V., Ananko, E. A., Podkolodnaya, O. A., Kolpakov, F. A., Podkolodny, N. L., and Kolchanov, N. A. (1998). Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. *Nucleic Acids Res* 26, 362-7.
- 10    Hirsch, D., Stahl, A., and Lodish, H. F. (1998). A family of fatty acid transporters conserved from mycobacterium to man. *Proc Natl Acad Sci U S A* 95, 8625-9.
- Hua, X., Liu, X., Ansari, D. O., and Lodish, H. F. (1998). Synergistic cooperation of TFE3 and smad proteins in TGF-beta-induced transcription of the plasminogen activator inhibitor-1 gene. *Genes Dev* 12, 3084-95.
- 15    Hui, T. Y., Frohnert, B. I., Smith, A. J., Schaffer, J. E., and Bernlohr, D. A. (1998). Characterization of the murine fatty acid transport protein gene and its insulin response sequence. *J Biol Chem* 273, 27420-9.
- Lai, E. (1992). Regulation of hepatic gene expression and development. *Semin Liver Dis* 12, 246-51.
- 20    Lee, Y. H., Yano, M., Liu, S. Y., Matsunaga, E., Johnson, P. F., and Gonzalez, F. J. (1994). A novel cis-acting element controlling the rat CYP2D5 gene and requiring cooperativity between C/EBP beta and an Sp1 factor. *Mol Cell Biol* 14, 1383-94.

- Martin, G., Schoonjans, K., Lefebvre, A. M., Staels, B., and Auwerx, J. (1997). Coordinate regulation of the expression of the fatty acid transport protein and acyl-CoA synthetase genes by PPARalpha and PPARgamma activators. *J Biol Chem* 272, 28210-7.
- 5 Memon, R. A., Fuller, J., Moser, A. H., Smith, P. J., Grunfeld, C., and Feingold, K. R. (1999). Regulation of putative fatty acid transporters and Acyl-CoA synthetase in liver and adipose tissue in ob/ob mice. *Diabetes* 48, 121-7.
- Motojima, K., Passilly, P., Peters, J. M., Gonzalez, F. J., and Latruffe, N. (1998). Expression of putative fatty acid transporter genes are regulated by peroxisome
- 10 proliferator-activated receptor alpha and gamma activators in a tissue- and inducer-specific manner. *J Biol Chem* 273, 16710-4.
- Rodenburg, R. J., Holthuizen, P. E., and Sussenbach, J. S. (1997). A functional Sp1 binding site is essential for the activity of the adult liver-specific human insulin-like growth factor II promoter. *Mol Endocrinol* 11, 237-50.
- 15 Rongneparut, P., Verdon, C. P., Gehrich, S. C., and Sul, H. S. (1991). Isolation and characterization of the transcriptionally regulated mouse liver (B-type) phosphofructokinase gene and its promoter. *J Biol Chem* 266, 8086-91.
- Ryu, S., Zhou, S., Ladurner, A. G., and Tjian, R. (1999). The transcriptional cofactor complex CRSP is required for activity of the enhancer-binding protein Sp1. *Nature*
- 20 397, 446-50.
- Schaffer, J., and Lodish, H. F. (1995). Molecular mechanisms of long-chain fatty acid



uptake. Trends in Cardiovascular Medicine 5, 218-224.

Schaffer, J. E., and Lodish, H. F. (1994). Expression cloning and characterization of a novel adipocyte long chain fatty acid transport protein [see comments]. Cell 79, 427-36.

- 5 Schoonjans, K., Staels, B., and Auwerx, J. (1996). The peroxisome proliferator activated receptors (PPARS) and their effects on lipid metabolism and adipocyte differentiation. Biochim Biophys Acta 1302, 93-109.

- Sorensen, P., and Wintersberger, E. (1999). Sp1 and NF-Y are necessary and sufficient for growth-dependent regulation of the hamster thymidine kinase promoter [In Process  
10 Citation]. J Biol Chem 274, 30943-9.

Stahl, A., Hirsch, D. J., Gimeno, R. E., Punreddy, S., Ge, P., Watson, N., Patel, S., Kotler, M., Raimondi, A., Tartaglia, L. A., and Lodish, H. F. (1999). Identification of the major intestinal fatty acid transport protein [In Process Citation]. Mol Cell 4, 299-308.

- 15 Stremmel, W. (1989). Mechanism of hepatic fatty acid uptake. Journal of Hepatology 9, 374-382.

All references cited herein are incorporated by reference in their entirety.

- While this invention has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled in the art that  
20 various changes in form and details may be made therein without departing from the spirit and scope of the invention as defined by the appended claims.

-115-

## CLAIMS

What is claimed is:

- 5           1.     An isolated nucleic acid comprising the nucleotide sequence of SEQ ID NO.:116 or its complement.
2.     An isolated nucleic acid comprising the coding sequence of SEQ ID NO.: 116.
3.     An isolated nucleic acid which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO.:117 or its complement.
- 10          4.     An isolated nucleic acid which hybridizes under stringency conditions of 6X SSC at 65° C, followed by at least two washes in 0.2X SSC/0.5% SDS at 65° C, to the nucleic acid comprising the nucleotide sequence of SEQ ID NO.:116.
- 15          5.     An isolated nucleic acid consisting of a nucleotide sequence having at least 95% identity to a nucleotide sequence of Claim 1.
6.     An isolated nucleic acid consisting of a nucleotide sequence having at least 90% identity to a nucleotide sequence of Claim 1.
- 20          7.     An isolated nucleic acid encoding a fusion polypeptide, wherein the isolated nucleic acid comprises a nucleotide sequence of SEQ ID NO.:116.

-116-

8. A vector comprising a nucleic acid of Claim 1.
9. A vector comprising a nucleic acid of Claim 2.
10. A vector comprising a nucleic acid of Claim 3.
11. A vector comprising a nucleic acid of Claim 4.
- 5 12. A vector comprising a nucleic acid of Claim 5.
13. A vector comprising a nucleic acid of Claim 6.
14. A vector comprising a nucleic acid of Claim 7.
15. An isolated host cell transfected with the vector of Claim 8.
16. An isolated host cell transfected with the vector of Claim 9.
- 10 17. An isolated host cell transfected with the vector of Claim 10.
18. An isolated host cell transfected with the vector of Claim 11.
19. An isolated host cell transfected with the vector of Claim 12.

20. An isolated host cell transfected with the vector of Claim 13.
21. An isolated host cell transfected with the vector of Claim 14.
22. A method of producing a polypeptide comprising the step of culturing  
the host cell of Claim 15 under conditions in which the nucleic acid is  
5 expressed, thereby producing the polypeptide.
23. A method of producing a polypeptide comprising the step of culturing  
the host cell of Claim 16 under conditions in which the nucleic acid is  
expressed, thereby producing the polypeptide.
24. A method of producing a polypeptide comprising the step of culturing  
10 the host cell of Claim 17 under conditions in which the nucleic acid is  
expressed, thereby producing the polypeptide.
25. A method of producing a polypeptide comprising the step of culturing  
the host cell of Claim 18 under conditions in which the nucleic acid is  
expressed, thereby producing the polypeptide.
26. A method of producing a polypeptide comprising the step of culturing  
15 the host cell of Claim 19 under conditions in which the nucleic acid is  
expressed, thereby producing the polypeptide.
27. A method of producing a polypeptide comprising the step of culturing  
the host cell of Claim 20 under conditions in which the nucleic acid is

expressed, thereby producing the polypeptide.

28. A method of producing a polypeptide comprising the step of culturing the host cell of Claim 21 under conditions in which the nucleic acid is expressed, thereby producing the polypeptide.
- 5 29. An isolated nucleic acid comprising at least 30 contiguous nucleotides of the nucleotide sequence of SEQ ID NO.:116.
30. An isolated nucleic acid comprising at least 200 contiguous nucleotides of the nucleotide sequence of SEQ ID NO.:116.
- 10 31. An isolated polypeptide comprising the amino acid sequence of SEQ ID NO.:117.
32. An isolated naturally occurring allelic variant of a polypeptide consisting of the amino acid sequence of Claim 31.
33. An isolated polypeptide consisting of an amino acid sequence having at least 95% identity to the amino acid sequence of Claim 31.
- 15 34. An isolated polypeptide consisting of an amino acid sequence having at least 90% identity to the amino acid sequence of Claim 31.
35. An isolated polypeptide encoded by a nucleic acid that hybridizes to a nucleic acid consisting of the nucleotide sequence of SEQ ID NO.:117

under stringency conditions of 6X SSC at 65° C, followed by at least two washes in 0.2X SSC/0.5% SDS at 65° C.

36. A fusion protein comprising a polypeptide consisting of the amino acid sequence of SEQ ID NO.:117.
- 5 37. The fusion protein of Claim 36, wherein the fusion protein transports fatty acids across a cell membrane or an artificial cell membrane system.
38. An isolated polypeptide comprising at least 15 contiguous amino acid residues of SEQ ID NO.:117.
- 10 39. An isolated polypeptide comprising at least 50 contiguous amino acid residues of SEQ ID NO.:117.
40. An isolated polypeptide comprising at least 360 contiguous amino acid residues of SEQ ID NO.:117.
- 15 41. An isolated polypeptide comprising an amino acid sequence having at least 15 contiguous amino acid residues of SEQ ID NO.:117, wherein the isolated polypeptide transports fatty acids across a cell membrane or an artificial cell membrane.
- 20 42. An isolated polypeptide encoded by a nucleic acid that hybridizes to a nucleic acid consisting of the nucleotide sequence of SEQ ID NO.:116 under stringency conditions of 6X SSC at 65° C, followed by at least

-120-

two washes in 0.2X SSC/0.5% SDS at 65° C.

- 5                   43.    A method for identifying an agent which binds to a protein comprising an amino acid sequence of SEQ ID NO.:117 comprising the steps of contacting the agent with the isolated protein under conditions appropriate for binding of the agent to the isolated protein, and detecting a resulting agent-protein complex.
44.    An agent identified by the method of Claim 43.
- 10               45.    A method for identifying an agent which is an inhibitor of fatty acid uptake by a protein encoded by a polynucleotide comprising a nucleotide sequence which encodes a protein consisting of the amino acid sequence of SEQ ID NO.:117, comprising the steps of:
- a)     maintaining test cells expressing said polynucleotide in the presence of a fatty acid and an agent to be tested as an inhibitor of fatty acid uptake;
- 15               b)     measuring uptake of the fatty acid in the test cells; and
- c)     comparing uptake of the fatty acid in the test cells with uptake of the fatty acid in suitable control cells;
- 20               wherein lower uptake of the fatty acid in the test cells compared to uptake of the fatty acid in the control cells is indicative that the agent is an inhibitor of fatty acid uptake by said protein.
46.    An inhibitor of fatty acid uptake identified by the method of Claim 45.
47.    The method of Claim 45 further comprising the steps of:

-121-

- 5
- a) administering the agent to one or more test animals;
  - b) measuring exogenously supplied fatty acids in one or more samples of tissue or bodily fluid from said test animals;
  - c) measuring exogenously supplied fatty acids in one or more comparable samples of tissue or bodily fluid from suitable control animals;
  - d) comparing the fatty acids of b) with the fatty acids of c);

whereby, lower fatty acids in step b) than in step c) is indicative that the agent is an inhibitor of said protein.

10      48.      An inhibitor of fatty acid uptake identified by the method of Claim 47.

49.      The method of Claim 45, wherein the nucleotide sequence which encodes a protein consists of a nucleotide sequence with 95% identity to a nucleotide sequence which encodes the polypeptide with SEQ ID NO.: 117.

15      50.      A method for identifying an agent which is an inhibitor of a protein encoded by a polynucleotide comprising a nucleotide sequence which encodes a protein comprising the amino acid sequence in SEQ ID NO.: 117 comprising the steps of:

- 20
- (a) introducing into host cells one or more vectors comprising a polynucleotide expressing said protein;
  - (b) culturing a first aliquot of the host cells with fatty acid substrate of said protein and with an agent being tested as an inhibitor of said protein;
  - (c) culturing a second aliquot of the host cells with fatty acid



-122-

substrate of said protein;

- (d) measuring, in the first and second aliquots, uptake of the fatty acid substrate of the host cells;

5 wherein less uptake of the fatty acid substrate in the first aliquot compared to the second aliquot is indicative that the agent is an inhibitor of said protein.

51. An inhibitor of fatty acid uptake identified by the method of Claim 50.

52. The method of Claim 50 further comprising the steps of:

- 10 a) administering the agent to one or more test animals;
- b) measuring exogenously supplied fatty acids in one or more samples of tissue or bodily fluid from suitable control animals;
- c) measuring exogenously supplied fatty acids in one or more comparable samples of tissue or bodily fluid from the test animals; and
- 15 d) comparing the fatty acids of the control animals with the fatty acids of the test animals whereby, lower fatty acids in the control animals than in the test animals is indicative that the agent is an inhibitor of said protein.

20 53. A method for identifying an agent which binds to a protein comprising an amino acid sequence of SEQ ID NO.:117 comprising the steps of contacting the agent with the isolated protein under conditions appropriate for binding of the agent to the isolated protein, and detecting a resulting agent-protein complex.

54. A method for identifying an agent which inhibits interaction between an isolated protein comprising an amino acid sequence of SEQ ID NO.:117, and further comprising a ligand of said protein, comprising:

(a) combining:

- 5 (1) said isolated protein;
- (2) the ligand of said protein; and
- (3) a candidate agent to be assessed for its ability to inhibit interaction between said protein of (1) and the ligand of (2), under conditions appropriate for interaction
- 10 between the said protein of (1) and the ligand of (2);
- (b) determining the extent to which said protein of (1) and the ligand of (2) interact; and
- (c) comparing the extent determined in (b) with the extent to which interaction of said protein of (1) and the ligand of (2) occurs in
- 15 the absence of the candidate agent to be assessed and under the same conditions appropriate for interaction of said protein of (1) with the ligand of (2);

wherein if the extent to which interaction of said protein of (1) and the ligand of (2) occurs is less in the presence of the candidate agent than

20 in the absence of the candidate agent, the candidate agent is an agent which inhibits interaction between said protein and the ligand of said protein.

55. A method for detecting, in a sample of cells, a nucleic acid molecule consisting of a nucleotide sequence with at least 90% sequence identity to SEQ ID NO.:116, comprising:

- 25 a) purifying nucleic acid from the cells;
- b) hybridizing 1) purified nucleic acid from the cells to 2) purified nucleic

acid comprising SEQ ID NO.:116, under conditions that allow hybridization between 1) and 2) if the sequences of 1) and 2) have at least 90% sequence identity; and

- 5 c) detecting resulting hybrid nucleic acids in the hybridization; wherein, if hybrid nucleic acids are detected at a significant level compared to a suitable control hybridization, then a nucleic acid molecule comprising at least 90% sequence identity to SEQ ID NO: 116, has been detected.

10 56. A method for identifying (1) nucleic acid molecules in fixed cells which specifically interact with a (2) nucleic acid molecule comprising the nucleotide sequence in SEQ ID NO.:116, said method comprising the steps of:

- 15 a) adding to the fixed cells the nucleic acid molecule comprising a nucleotide sequence in SEQ ID NO.:116;
- b) incubating the fixed cells under conditions allowing hybridization of (1) with (2);
- c) removing the nucleic acid molecule of step a) that has not hybridized; and
- d) detecting hybrid molecules comprising (1) and (2).

20 57. A method for detecting FATP3 in a sample of cells, comprising the steps of adding an agent that specifically binds to FATP3 to the sample, and detecting the agent specifically bound to the FATP3.

58. The method of Claim 57 wherein the agent is an antibody which specifically binds to FATP3.

59. A method for detecting FATP3 in a sample of cell lysate, comprising the steps of adding an agent that specifically binds to FATP3 to the sample, and detecting agent specifically bound to the FATP3 .
- 5 60. The method of Claim 59 wherein the agent is an antibody which specifically binds to FATP3.
61. An isolated antibody which binds to a polypeptide having an amino acid sequence consisting of at least 95% amino acid sequence identity with the amino acid sequence of SEQ ID NO.:117.
- 10 62. An isolated antibody which binds to a fatty acid transport protein having the amino acid sequence of SEQ ID NO.:117.
63. A method for detecting, in a sample of cells, a nucleic acid molecule comprising at least 90% sequence identity to SEQ ID NO.:116 comprising:
- 15 — a) purifying nucleic acid from the cells;
- b) hybridizing 1) purified nucleic acid from the cells to 2) purified nucleic acid comprising SEQ ID NO.:116 under conditions that allow hybridization between 1) and 2) if the sequences of 1) and 2) have at least 90% sequence identity; and
- 20 c) detecting resulting hybrid nucleic acids in the hybridization; wherein, if hybrid nucleic acids are detected at a significant level compared to a suitable control hybridization, then a nucleic acid molecule having at least 90% sequence identity to SEQ ID NO.:116 has been detected.

64. A method for detecting, in a sample of purified nucleic acid, a nucleic acid molecule comprising at least 90% sequence identity to SEQ ID NO.: 116 comprising:
- 5           a)     hybridizing 1) the sample of purified nucleic acid to 2) purified nucleic acid comprising SEQ ID NO.:116 under conditions that allow hybridization between 1) and 2) if the sequences of 1) and 2) have at least 90% sequence identity; and
- 10           b)     detecting resulting hybrid nucleic acids in the hybridization; wherein, if hybrid nucleic acids are detected at a significant level compared to a suitable control hybridization, then a nucleic acid molecule having at least 90% sequence identity to SEQ ID NO.:116 has been detected.
65. A method for detecting FATP3 in a sample of cells, comprising the steps of adding an agent that specifically binds to FATP3 to the
- 15           sample, and detecting agent specifically bound to the FATP3.
66. The method of Claim 65 wherein the agent is an antibody which binds to FATP3.
67. A vector comprising a FATP regulatory sequence and at least one targeting sequence directed to the regulatory region of a nucleic acid
- 20           with a nucleotide sequence selected from the group consisting of:
- a)     SEQ ID NO.:46
- b)     SEQ ID NO.:48
- c)     SEQ ID NO.:116
- d)     SEQ ID NO.:52

- e) SEQ ID NO.:54 and
- f) SEQ ID NO.:56

68. An isolated host cell transfected with a vector of Claim 67.
- 5 69. A method of producing a polypeptide comprising culturing the host cell of Claim 68 under conditions in which the nucleic acid is expressed, thereby producing the polypeptide.
70. An isolated nucleic acid comprising a nucleotide sequence encoding a functional portion or fragment of a FATP polypeptide comprising a lipocalin domain.
- 10 71. The isolated nucleic acid of Claim 70 further comprising a nucleotide sequence encoding upstream amino acid residues.
- 15 72. An isolated nucleic acid comprising a nucleotide sequence encoding a portion or fragment of a FATP protein containing a lipocalin domain, wherein the nucleotide sequence is selected from the group consisting of portions or fragments of:
- a) SEQ ID NO.:46
  - b) SEQ ID NO.:48
  - c) SEQ ID NO.:116
  - d) SEQ ID NO.:52
  - 20 e) SEQ ID NO.:54 and
  - f) SEQ ID NO.:56.

73. An isolated nucleic acid of Claim 72 further comprising at least about 90 nucleotides of the sequence upstream of the lipocalin domain.
74. A vector comprising a nucleic acid of Claim 73.
75. An isolated host cell comprising the vector of Claim 74.
- 5 76. A method of producing a polypeptide comprising the step of culturing the host cell of Claim 75 under conditions in which the nucleic acid is expressed, thereby producing the polypeptide.
77. A functional portion or fragment of a FATP polypeptide comprising a lipocalin domain.
- 10 78. The FATP polypeptide of Claim 77 further comprising upstream amino acid residues.
79. An isolated polypeptide comprising an amino acid sequence containing a FATP lipocalin domain, wherein the amino acid sequence is selected from the group consisting of portions or fragments of:
- 15 a) SEQ ID NO.:47;  
b) SEQ ID NO.:49;  
c) SEQ ID NO.:117;  
d) SEQ ID NO.:53;  
e) SEQ ID NO.:55; and  
20 f) SEQ ID NO.:57.

- 5                   80.    A functional portion or fragment of a FATP polypeptide comprising an amino acid sequence selected from the group consisting of:
- a)     SEQ ID NO.:126;
  - b)     SEQ ID NO.:127;
  - c)     SEQ ID NO.:128;
  - d)     SEQ ID NO.:129;
  - e)     SEQ ID NO.:130; and
  - f)     SEQ ID NO.:131.
- 10               81.    A fusion protein comprising a polypeptide consisting of a FATP polypeptide containing a lipocalin domain.
82.    The fusion protein of Claim 81 further comprising upstream sequences.
83.    The fusion protein of Claim 82, wherein the upstream sequences comprise at least about 30 amino acid residues of an upstream sequence.
- 15               84.    A fusion protein comprising a polypeptide consisting of a FATP polypeptide containing a lipocalin domain, wherein the polypeptide consists of an amino acid sequence selected from the group consisting of portions or fragments of:
- a)     SEQ ID NO.:47;
  - b)     SEQ ID NO.:49;
  - c)     SEQ ID NO.:117;
  - d)     SEQ ID NO.:53;
- 20



-130-

- e) SEQ ID NO.:55; and
- f) SEQ ID NO.:57.

85. The fusion protein of Claim 84 further comprising upstream sequences.
- 5 86. A method for identifying an agent which binds to a polypeptide, wherein the polypeptide comprises a FATP lipocalin domain, comprising the steps of contacting the agent with the polypeptide under conditions appropriate for binding of the agent to the polypeptide, and detecting a resulting agent-polypeptide complex.
87. The agent identified by the method of Claim 86.
- 10 88. A method for identifying an agent which binds to a polypeptide, wherein the polypeptide comprises a FATP lipocalin domain and about 30 amino acid residues of an upstream sequence, comprising the steps of contacting the agent with the polypeptide under conditions appropriate for binding of the agent to the polypeptide, and detecting a
- 15 resulting agent-polypeptide complex.
89. The agent identified by the method of Claim 88.
90. A method for identifying an agent which binds to a polypeptide, wherein the polypeptide comprises a FATP lipocalin domain and consists of an amino acid sequence selected from the group consisting
- 20 of portions or fragments of:
- a) SEQ ID NO.:47;

-131-

- b) SEQ ID NO.:49;
- c) SEQ ID NO.:117;
- d) SEQ ID NO.:53;
- e) SEQ ID NO.:55; and
- 5 f) SEQ ID NO.:57,

comprising the steps of contacting the agent with the polypeptide under conditions appropriate for binding of the agent to the polypeptide, and detecting a resulting agent-polypeptide complex.

91. An agent identified by the method of Claim 90.

10 92. A method for identifying an agent which binds to a polypeptide, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of:

- a) SEQ ID NO.:126;
- b) SEQ ID NO.:127;
- 15 c) SEQ ID NO.:128;
- d) SEQ ID NO.:129;
- e) SEQ ID NO.:130; and
- f) SEQ ID NO.:131,

20 comprising the steps of contacting the agent with the polypeptide under conditions appropriate for binding of the agent to the polypeptide, and detecting a resulting agent-polypeptide complex.

93. An agent identified by the method of Claim 92.

-132-

94. A method for identifying an agent which binds to a polypeptide comprising a FATP lipocalin domain, wherein the polypeptide is encoded by a nucleotide sequence consisting of portions or fragments of:

- 5           a)     SEQ ID NO.:46;  
          b)     SEQ ID NO.:48;  
          c)     SEQ ID NO.:116;  
          d)     SEQ ID NO.:52;  
          e)     SEQ ID NO.:54; and  
10          f)     SEQ ID NO.:56.

comprising the steps of contacting the agent with the polypeptide under conditions appropriate for binding of the agent to the polypeptide, and detecting a resulting agent-polypeptide complex.

95. An agent identified by the method of Claim 94.

- 15          96. A method for identifying an agent which binds to a polypeptide comprising a FATP lipocalin domain and upstream sequences, wherein the polypeptide is encoded by a nucleotide sequence consisting of portions or fragments of:

- a.     SEQ ID NO.:46;  
20          b.     SEQ ID NO.:48;  
          c.     SEQ ID NO.:116;  
          d.     SEQ ID NO.:52;  
          e.     SEQ ID NO.:54; and  
          f.     SEQ ID NO.:56.

-133-

comprising the steps of contacting the agent with the polypeptide under conditions appropriate for binding of the agent to the polypeptide, and detecting a resulting agent-polypeptide complex.

97. An agent identified by the method of Claim 96.
- 5 98. An isolated nucleic acid sequence comprising the nucleic acid sequence of SEQ ID NO: 113.
99. The portion of the isolated nucleic acid sequence of Claim 98 which encodes a FATP regulatory protein.
- 10 100. The portion of the isolated nucleic acid sequence of Claim 98 which encodes a FATP5 promoter.
101. A method of identifying an agent which alters the level of expression of the nucleic acid encoding an FATP protein comprising:  
determining a base level of expression of the nucleic acid encoding the FATP protein;
- 15 (b) contacting an agent with an isolated nucleic acid containing the coding region of the FATP protein under functional control of its promoter under conditions suitable for binding of the agent to the promoter;
- (c) maintaining agent-promoter binding during expression of the FATP protein; and
- 20 (d) comparing the level of expression of the agent bound promoter to that of the baseline level of expression,

whereby, if the level of expression of the agent bound promoter is significantly different from that of the baseline level of expression, then an agent which alters the level of expression of the nucleic acid encoding the FATP protein has been identified.

- 5           102.   The method of Claim 101, wherein the FATP protein is FATP2.
103.   The method of Claim 102, wherein the FATP2 is encoded by a nucleic acid comprising the nucleotide sequence of SEQ ID NO: 48.
104.   The method of Claim 102, wherein the FATP2 comprises the amino acid sequence of SEQ ID NO: 49.
- 10          105.   The method of Claim 102, wherein expression is inhibited.
106.   The method of Claim 102, wherein expression is promoted.
107.   The method of Claim 101, wherein the FATP protein is FATP5.
108.   The method of Claim 107, wherein the FATP5 is encoded by a nucleic acid comprising the nucleotide sequence of SEQ ID NO: 54.
- 15          109.   The method of Claim 107, wherein the FATP5 comprises the amino acid sequence of SEQ ID NO: 55.
110.   The method of Claim 107, wherein expression is inhibited.

111. The method of Claim 107, wherein expression is promoted.
112. A method for directing an agent to liver cells in a mammal, comprising administering to the mammal a complex which comprises the agent and a moiety which binds to FATP2.
- 5 113. The method of Claim 112, wherein the agent alters fatty acid uptake in liver cells.
114. The method of Claim 112, wherein the agent alters the level of fatty acids in bile.
- 10 115. A method for directing an agent to the gall bladder in a mammal, comprising administering to the mammal a complex which comprises the agent and a moiety which binds to FATP2.
116. The method of Claim 115, wherein the agent alters the level of fatty acids in bile.
- 15 117. A method for directing an agent to the liver in a mammal, comprising administering to the mammal a complex which comprises the substance and a moiety which binds to FATP5.
118. The method of Claim 117, wherein the agent alters the uptake of fatty acids in liver cells.

119. The use of an isolated nucleic acid comprising the nucleotide sequence of SEQ ID NO.:116 or its complement in the manufacture of a medicament.
- 5 120. The use of an isolated polypeptide comprising the amino acid sequence of SEQ ID NO.:117 in the manufacture of a medicament.
121. The use of an agent which is an inhibitor of fatty acid uptake of a protein with the amino acid sequence of SEQ ID NO.:117 in the manufacture of a medicament.
- 10 122. The use of an isolated nucleic acid comprising a nucleotide sequence encoding a portion or fragment of a FATP protein containing a lipocalin domain in the manufacture of a medicament, wherein the nucleotide sequence is selected from the group consisting of portions or fragments of:
- 15 a) SEQ ID NO.:46  
b) SEQ ID NO.:48  
c) SEQ ID NO.:116  
d) SEQ ID NO.:52  
e) SEQ ID NO.:54 and  
f) SEQ ID NO.:56.
- 20 123. The use of an isolated polypeptide comprising an amino acid sequence containing a FATP lipocalin domain in the manufacture of a medicament, wherein the amino acid sequence is selected from the

-137-

group consisting of portions or fragments of:

- 5
- a) SEQ ID NO.:47;
  - b) SEQ ID NO.:49;
  - c) SEQ ID NO.:117;
  - d) SEQ ID NO.:53;
  - e) SEQ ID NO.:55; and
  - f) SEQ ID NO.:57.

10 124. The use of an isolated polypeptide in the manufacture of a medicament, the polypeptide comprising an amino acid sequence selected from the group consisting of:

- 15
- 1. SEQ ID NO.:126;
  - 2. SEQ ID NO.:127;
  - 3. SEQ ID NO.:128;
  - 4. SEQ ID NO.:129;
  - 5. SEQ ID NO.:130; and
  - 6. SEQ ID NO.:131.

20 125. The use of an isolated polypeptide in the manufacture of a medicament for treating obesity, the polypeptide comprising an amino acid sequence selected from the group consisting of:

- 7. SEQ ID NO.:126;
- 8. SEQ ID NO.:127;
- 9. SEQ ID NO.:128;
- 10. SEQ ID NO.:129;



-138-

11. SEQ ID NO.:130; and
12. SEQ ID NO.:131.

[illegible]

W	S	A	A	A	F	C	V	Y	V	G	G	G	W	R	F	L	R	I	V	C	K	T	A	R	R	D	L	F	Q	L	S	V	L	I	R			
V	L	A	G	L	A	G	-	L	L	L	L	P	L	L	L	T	C	C	C	P	Y	L	L	Q	D	V	R	Y	F	L	R	L	A	N	M	A		
S	S	E	S	G	C	S	L	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-				
L	S	L	V	G	A	A	L	T	L	F	L	L	P	L	Q	P	P	P	G	L	R	W	L	H	K	D	V	A	F	T	F	K	M	L	F	Y	G	
V	I	L	A	G	V	L	L	Y	I	T	V	V	H	G	D	F	I	Y	R	S	Y	L	T	L	N	R	D	L	T	G	L	A	L	I	L	E	S	
I	T	P	I	Q	K	S	L	G	Y	L	F	G	N	Y	F	D	E	L	D	D	L	A	T	R	M	P	R	V	L	A	D	T	P	V	I	V	R	G
S	D	Y	Y	G	G	A	H	T	T	V	-	-	-	-	-	-	L	I	D	D	L	A	T	R	V	L	A	D	T	P	V	I	V	R	G	-	-	

**SUBSTITUTE SHEET (RULE 26)**

FIG. 1D

[illegible]

mmFATP1	126	- - - - -	A	P	G	D	V	V	A	V	F	L	E	G	R	P	E	F	V	G	L	W	L	G	L	A	K	A	G	V	V	A	A	L	L	N			
mmFATP2	101	- - - - -	L	R	Q	G	D	C	V	A	L	F	M	G	N	E	P	A	Y	V	W	I	W	L	G	L	L	K	L	G	C	P	M	A	C	L	N		
mmFATP3	94	- - - - -	L	A	P	G	A	T	V	A	L	L	L	P	A	G	P	D	E	L	W	I	W	F	G	L	A	K	A	G	L	R	T	A	F	V	P		
mmFATP4	8	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -			
mmFATP5	140	- - - - -	I	Q	N	T	R	D	A	A	I	L	V	L	P	S	K	T	I	S	A	L	S	V	F	L	G	L	A	K	L	G	C	P	V	A	L	I	N
ceFATPa	125	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -		
scFATP	134	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -		
mtFATP	94	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -	- - - - -		

FIG. 1E

V	N	L	R	R	E	P	L	A	F	C	L	G	T	S	A	A	K	A	L	I	V	G	G	E	M	A	A	A	V	A	E	V	S	E		
Y	N	I	R	A	K	S	L	L	H	C	F	Q	C	C	G	A	K	V	L	L	A	S	P	P	D	L	Q	E	A	V	E	E	V	L	P	
T	A	L	R	R	Q	P	L	L	H	C	L	R	S	C	G	A	S	A	L	V	L	A	T	E	F	F	L	E	S	L	E	P	D	L	P	
T	N	L	R	R	D	A	L	L	R	H	C	L	S	T	S	K	A	R	A	L	I	F	G	S	E	E	M	A	S	A	I	C	E	I	H	A
P	H	S	R	Q	M	P	L	L	H	S	V	R	S	S	G	A	S	V	L	I	V	D	P	P	D	L	Q	E	N	L	E	E	V	L	P	
S	N	L	K	R	E	Q	L	L	V	H	C	I	T	A	S	K	T	K	A	I	L	T	S	V	T	L	L	Q	N	I	M	L	D	A	I	D
Y	N	T	K	G	T	P	L	L	V	H	S	L	K	I	S	N	I	T	Q	V	F	I	D	P	D	A	S	N	P	I	R	E	S	E	E	
Y	H	Q	R	G	E	V	L	A	H	S	L	G	L	L	D	A	K	V	L	L	A	E	S	D	L	V	S	A	V	A	E	E	C	G	A	

FIG. 1F

mmFATP1	195	Q L Q	K S L L	K F C S G	D L	- G P E S	L P D T Q	L L D P M L A E A P	T T P L
mmFATP2	171	T L L	- - K D A	V S V F	Y V	- S R T S	N T N G V D	T I L D K V D G V S A	E P T
mmFATP3	164	A L L	- - A M G	L H L W	A T	- G P E T	N V A G I S	L L S E A A D Q V D	E P V
mmFATP4	56	S L L	- - P T L	S G S G	S W	- E P S T	V P V S T	L L S E A A D Q V D	E P V
mmFATP5	213	K L L	- - A E N	I H C F	Y L	- G H S S	P T P G V E	L L S E A A D Q V D	E P V
ceFATPa	194	O K L	- - F D V	E G I E	V S V	- G E P K	N S G F K N	L L S E A A D Q V D	E P V
scFATP	204	E I K	- - N A L	P D V K	L N Y L	- E E Q D	L M H E L N	S Q S P E F L Q	D N V R T
ntFATP	164	S R Q	- - - -	- - R V A	- -	- G D V	L T V E	L L S E A A D Q V D	E P V

FIG. 1G

A Q A P	G K	- -	G M D	D R L F	Y I Y T	S G T	T G L P	K A A	I V V H S
P E S W	R S A	- -	T T F	T P A V	Y I Y T	S G T	T G L P	K A A	T I N H H
P G Y L	S A	- -	T T M	D T C L	Y I Y T	S G T	T G L P	K A A	I S H L
P S H P	D K	- -	T T F	D K L F	Y I Y T	S G T	T G L P	K A A	I V V H S
P A S L	R A T	- -	I K W	K S P A	I F I F	T S G T	T G L P	K A A	I L S H E
K T L D	I V -	- -	D F K	S I L C	F I Y T	S G T	T G M P	K A A	V M K H F
P L G L	T D -	- -	- - F	- - K	P S M L	I Y T	S G T	T G L P	K S A I M S W R
P A S A	S A -	- -	V Q A	K D T A	F Y I F	T S G T	T G F P	K A A	S V M T H

FIG. 1H

mmFATP1	265	R	V	Y	R	I	A	A	F	G	H	S	Y	S	M	R	A	-	-	A	D	V	L	Y	D	C	L	P	L	Y	H	S	A	G	N	
mmFATP2	241	R	L	R	Y	Q	T	G	-	L	A	M	S	S	G	I	T	A	-	-	Q	D	V	I	Y	T	T	M	P	L	Y	H	S	A	L	
mmFATP3	234	K	V	L	Q	C	Q	Q	-	F	Y	H	L	C	G	V	H	Q	-	-	E	D	V	I	Y	L	A	L	P	L	Y	H	S	A	L	
mmFATP4	125	R	Y	Y	R	M	A	S	-	L	V	Y	Y	G	F	R	R	P	-	-	E	D	V	I	Y	D	C	L	P	L	Y	H	S	R	K	
mmFATP5	283	R	V	I	Q	V	S	N	-	V	L	S	F	C	G	C	R	A	-	-	D	D	V	Y	V	D	V	L	P	L	Y	H	T	I	G	L
ceFATPa	264	R	Y	Y	S	I	A	V	G	A	A	K	S	F	G	I	R	P	-	-	S	D	R	M	Y	V	S	M	P	L	Y	H	T	A	A	Q
scFATP	273	K	S	S	V	G	C	Q	V	F	G	H	V	L	H	M	T	N	-	-	E	S	T	Y	F	T	A	M	P	L	Y	H	S	T	A	A
mtFATP	223	R	W	L	R	A	L	A	-	V	F	G	G	M	G	L	R	L	K	G	S	D	T	L	Y	S	C	L	P	L	Y	H	N	N	A	L

FIG. 1I

I	M	G	V	G	O	C	V	I	V	G	L	T	V	V	L	R	K	K	F	S	A	S	R	F	W	D	D	C	V	K	Y	N	C	T	V	V	Q		
M	I	G	L	H	G	C	I	V	V	G	A	X	X	X	L	C	D	K	F	S	A	S	Q	F	W	D	D	C	R	K	Y	N	V	T	V	I	Q		
L	L	G	I	V	G	C	L	G	I	G	A	T	V	V	L	K	P	K	F	S	A	S	Q	F	W	D	D	C	Q	K	H	R	V	T	V	F	Q		
H	R	G	D	W	Q	C	L	L	H	G	M	T	V	V	I	R	K	K	F	S	A	S	R	F	W	D	D	C	I	K	Y	N	C	T	V	V	Q		
V	L	G	F	L	G	C	L	Q	V	G	A	T	C	V	L	A	P	K	F	S	A	S	R	F	W	D	D	C	R	O	H	G	V	T	V	I	L		
I	L	G	V	G	O	A	L	L	G	G	S	S	C	V	I	R	K	K	F	S	A	S	N	F	W	D	D	C	V	K	Y	D	C	T	V	S	O		
L	L	G	A	C	A	I	L	S	H	G	Q	C	L	A	L	S	H	K	F	S	A	S	T	F	W	D	D	C	Q	V	Y	L	T	G	A	T	H	I	Q
T	V	A	V	S	S	V	I	N	S	G	A	T	L	A	L	G	K	S	F	S	A	S	R	F	W	D	D	C	V	I	A	N	R	A	T	A	F	V	

FIG. 1J

336  
3311  
304  
196  
353  
335  
344  
295

Y	I	G	E	I	C	R	Y	L	L	R	Q	P	V	R	D	V	E	Q	R	H	R	V	R	L	A	V	C	N	G	L	R	P	A
Y	I	G	E	L	L	C	R	Y	L	C	T	P	Q	K	P	N	D	R	D	H	K	V	K	K	A	L	C	N	G	L	R	Q	D
Y	I	G	E	L	C	R	Y	L	L	N	O	P	P	S	K	A	E	F	D	H	K	V	R	L	A	V	C	N	G	L	R	P	D
Y	I	G	E	L	C	R	Y	L	L	N	O	P	P	R	E	A	E	S	R	H	K	V	R	M	A	L	C	N	G	L	R	Q	S
Y	V	G	E	I	C	R	Y	L	L	N	Y	P	T	Q	P	E	D	K	L	H	T	V	R	L	A	M	C	T	G	L	R	A	N
Y	I	G	E	I	C	R	Y	L	L	A	Q	P	V	V	E	E	E	S	R	H	R	M	V	L	L	V	C	N	G	L	R	A	E
Y	V	G	E	V	C	R	Y	L	L	H	T	P	I	S	K	Y	E	K	M	H	X	V	K	V	A	Y	C	N	G	L	R	P	E
Y	I	G	E	I	C	R	Y	L	L	N	O	P	P	A	K	P	T	R	A	H	Q	V	R	V	I	C	N	G	L	R	P	E	E

FIG. 1K

[illegible]

FIG. 11

mmFATP1	406	CGFN	SRILLTH	VYPIRLLVKVNEDTMEPL
mmFATP2	381	VQRA	NYLQRK	VARVELLIKXVDVETKDEPVI
mmFATP3	374	VQRA	SWLYKH	IFPFSLLIRYDVMTQEPIL
mmFATP4	266	CGFN	SRILLSF	VYPIRLLVKVNEDTMEPLI
mmFATP5	423	VQRT	SCILLRM	LTPPFEELVQFDIEETAEPL
ceFATPa	404	CQFLPG	SPLLTKK	MHPVRLIKVDDVETGFAI
scFATP	417	CRNYG	TIILQWF	SFQQQLLVRRMDDPNDDSVI
mtFATP	365	AGVVS	-----	PMPRLAFVEYDLDLDTGDPPL

FIG. 1M

-	RDSEGL	CI	PCQP	GE	PG	LL	LV	QG	Q	IN	Q	Q	DP	PL	RR	FF	DD	GG	YV	-	SS	AT
-	RDAN	GY	CI	KVPK	GE	VL	LV	CV	CK	IT	OL	TP	PP	FI	GG	YA	AG	-	-	-	GT	QT
-	RNA	OGH	CM	TS	PG	QL	LV	VA	CP	VS	OO	SP	PF	FL	GG	YA	AG	-	-	-	AP	EL
-	RGP	DD	CI	PCQP	GG	QL	LV	QG	RR	IR	Q	Q	DP	PL	RR	FF	DD	GG	YL	-	NG	AN
-	RDK	OGF	CI	PE	PG	QL	LV	VT	KT	VR	X	N	Q	PP	FL	GG	YR	GG	-	-	QA	ES
-	RTS	DL	CI	AC	NP	GM	V	ST	TR	IR	X	N	NP	PL	LL	QS	FF	GG	YL	-	SK	ET
-	RRNS	KGL	CE	VAP	VG	EM	LS	MR	IF	FF	P	K	K	PP	ET	SS	FF	GG	YL	-	NA	KE
-	RDAS	GR	VR	RR	VD	GL	LS	SR	VN	-	R	L	Q	PP	FF	DD	GG	YT	-	-	PV	AS

FIG. 1N



FIG. 10

FIG. 1P

544  
544  
517  
510  
404  
559  
544  
562  
494

V V Y N E W S L H  
V K V E E K R R G H  
E L I A S K R S L S  
E L H H V L V L L V  
M Y Q H S T H R L T V  
L F Q Q A R D T V  
L Y Q V G A R R L A

**9/202**

Q	K	T	R	L	Q	R	E	G	F	D
R	K	V	T	L	M	E	E	G	F	D
O	K	V	R	L	A	N	E	G	F	D
Q	K	T	E	L	R	K	E	G	F	D
V	K	S	R	L	V	R	E	G	F	D
V	K	T	N	L	Q	R	L	G	I	M
I	K	F	-	-	-	-	-	-	-	-
R	K	V	E	L	R	N	Q	A	Y	G

mmFATP1	611	P	R	Q	T	S	D	R	L	F	F	L	D	L	K	Q	G	R	Y	V	P	L	D	E	R	V	H	A	R	I	C	A	Q	D	F	S	L
mmFATP2	585	P	T	V	I	K	D	D	L	Y	F	M	D	D	A	E	K	T	F	V	P	M	T	E	N	I	Y	N	A	I	I	D	K	T	L	K	
mmFATP3	578	P	S	V	L	S	D	D	L	V	V	L	D	D	D	I	G	A	Y	V	P	L	T	P	A	R	Y	S	A	L	L	S	G	D	L		
mmFATP4	471	P	S	V	V	K	D	D	L	F	Y	L	D	D	R	K	G	C	Y	V	A	L	D	Q	E	A	Y	T	R	I	Q	A	G	E	K		
mmFATP5	627	V	G	I	I	A	D	P	L	Y	I	L	D	D	A	Q	T	F	R	S	L	M	P	D	V	Y	Q	A	V	C	E	G	T	W			
ceFATPa	616	D	-	A	P	S	D	S	I	Y	I	Y	N	S	E	N	R	N	F	V	P	F	R	N	D	L	R	C	K	V	S	L	Q	S			
scFATP		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-				
mtFATP	562	A	-	D	I	E	D	P	L	Y	V	L	A	G	P	D	E	G	Y	V	P	Y	Y	A	E	Y	P	E	E	V	S	L	Q	R			

FIG. 1S

FIG. 2A

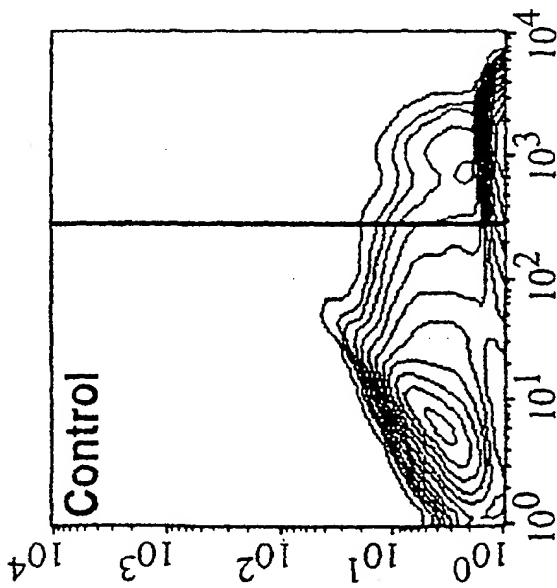


FIG. 2B

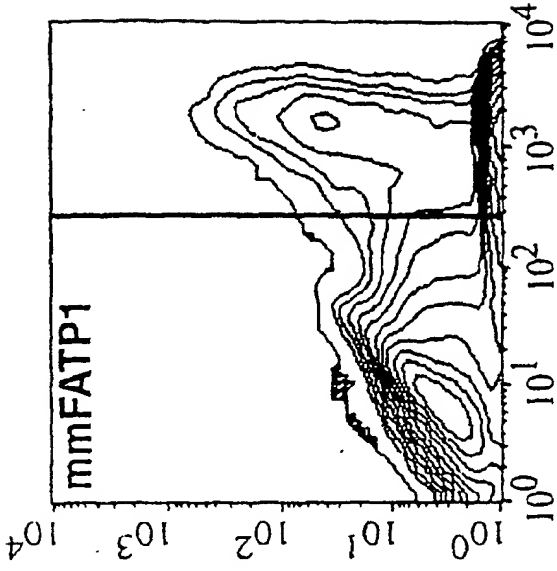


FIG. 2C

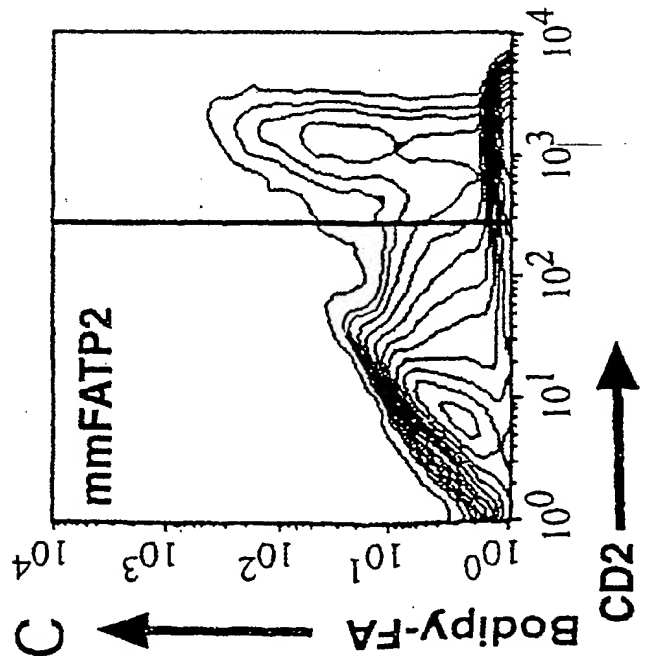


FIG. 2D

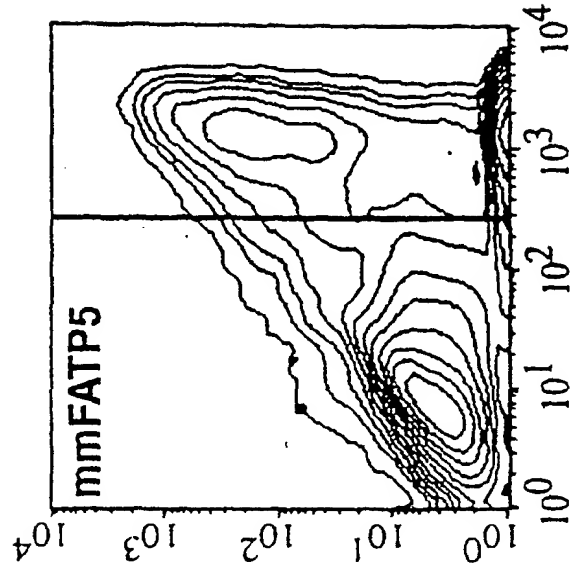


FIG. 3

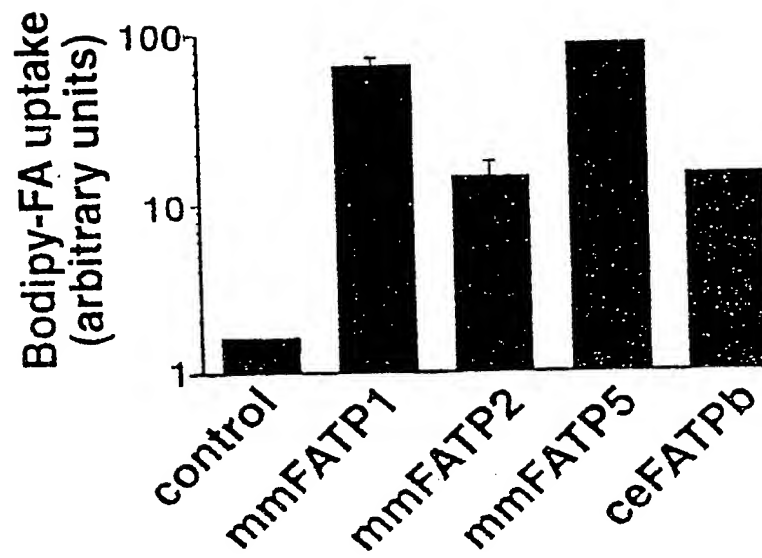


FIG. 4

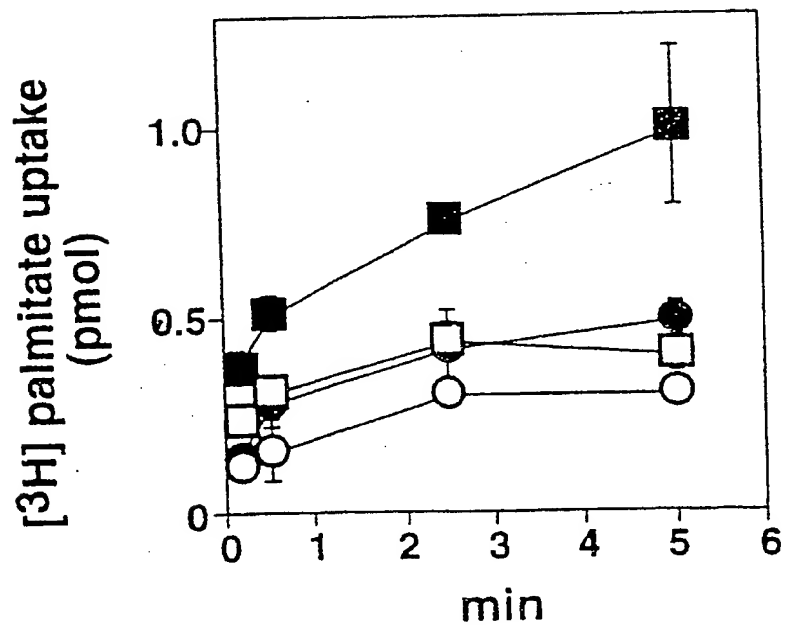
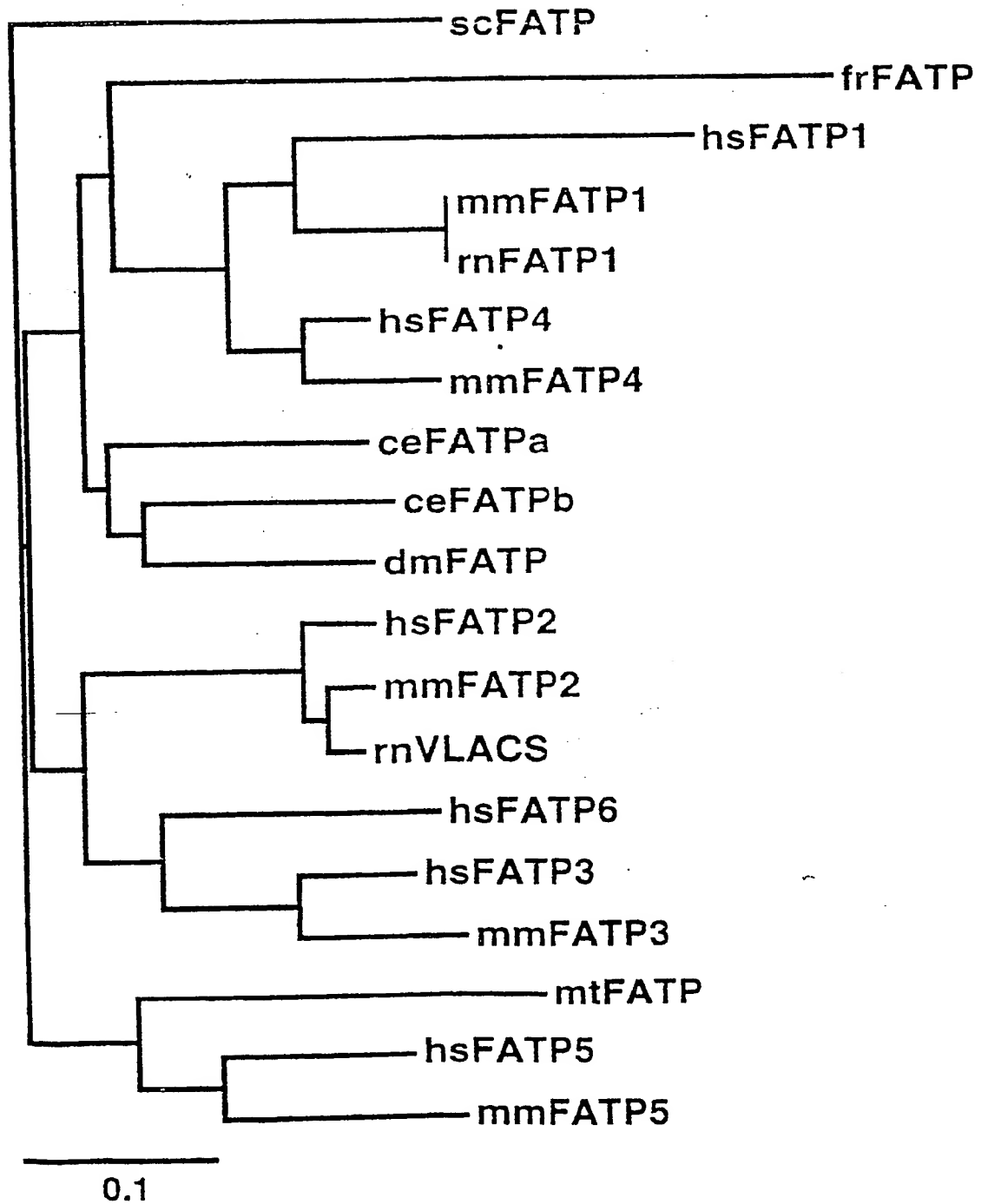


FIG. 5



**SUBSTITUTE SHEET (RULE 26)**

FIG. 6

[illegible]

FIG. 7



mmFATP3 DNA sequence

```
ACGACTCACTATAGGCAACAGCTATCAAGTGGCATGCAC 40
CGCTAAGCTTGGGCCCCCTCGACGGATCTCTATACAGGGCC 80
GCGCAACCCGCAAGGCTCTCAACAGCGCGTGCAGTCTGGGCT 120
GGCGTCTGGCGTACCTTGGCCCCCGGCAAGAGCCCAACACAC 160
CTTCTTCATCGACGGCGCGCGACGGCTTTAGCTACGGCGAC 200
GCTCAAGCGCAACAGCAACCGCAATTCTCTGGCGCTTTCTTC 240
CGCGACGGCGCGTGCACCGCGCGCGCGCGCAAGCTTGGCGAC 280
CGCGACGACTGACCAAGCGCGACGGCTGGCGCGCTCCGGCT 320
CGCACTTCGGCTCGTACAGGCAAGAGCGCGCGCGCGCTCTTC 360
CAACCGCGCGCGCAACGTGGCGCTGCTCTCTCCACGGCGCC 400
```

FIG. 8A

GCATTTTCCTTTGCAATTTTCGTTCCGACTGGCCAA-----  
 CTCGGCAGCGCCCTTTTGTGCCCCACCGCTTTAAGCCGAGGAC 480  
 CCGTGGTGCACCTGGCTCCGACGCTGGGGTCCAGTGGCGCT 520  
 CGTGGTGGCCACACAGTTTCTGCACTCCCTGGCAGCCCGAC 560  
 CTCGGCGGCTTTGACAGCCATGGGGCTCCACCTATGGGCGA 600  
 CCGGCGCTGAAACTAATGTAGCTGCAATCAGCAATTTTGGT 640  
 ATCGCAAGCAGCACAACCAAGTGCATCAGCCAGTGGCGGG 680  
 TACCTCTCTGCCCCCCACAACATAATGGAACCTGCGCTGT 720  
 ACATCTTCACTCTGGCACTACCTGGCTGGCCAGGCTGC 760  
 TCGAATCAGTCACTCTCAAGGTTCTACAGTGGCAGGCAATC 800  
 TACCATCTGTGTGGAGTCCACACGAGGAGGAGGATCTACC 840  
 TCGCACTCCCACTGTACACACATGTCTGGCTCCCTTCTGGG 880  
 CATTCGTGGGCTGCTTGGGCAATGGGGCCACCGTGGTGGT 920  
 AAAACCAAGTTCTCAGCTAGCCAGTTCTGGCAAGATTGCG 960  
 AGAACAACAGGGTGCACAGTGTCCAGTACATTTGGGAGTT 1000  
 CTGCGCATACCTGGTCAACAGCCCCCGACCAAGGCAAC 1040  
 TTTTCAACATAAGGTGGGCTTGGCAGTGGGCAAGTGGCTGC 1080  
 GCGCACAACCTGGCAGCGTTTCTGGGGCAATTTGCAAC 1120  
 TCTGCACATACTGCAACGATGGCATCACAAGGGCAAC 1160  
 GTACCTACCTTCAATTACACAGCAACCGCAGCGTGCAGTGG 1200  
 GCGCAGCTTCTGGCTTTACCAAGCACATCTTCCCTTCTC 1240  
 CTTCATTCGATACCTGTCTCATCACAAGGGCAAGCTATTGG 1280  
 AATGCCCCAGGGGCACTGCATCAACCAATCTCCAGGTGAG 1320  
 CAGGCTTACGTGGTGGCCCCAGTCAAGCAGTACGCCCCCT 1360  
 CCTGGGCTATGCTGGGGCTCCGCAAGCTGGCAAGGCAAC 1400  
 CTGCTCAAGCATGTCTCTGGTCTGGGCAAGTTTCTCTCA 1440  
 ATACTGGGCAAGCTCTGGTCTGTGATCAGCAAGGCTTTCT 1480  
 TCACCTCCACCATCTGCTGCAACCAATCAGGTGCAAG 1520  
 GCAACAATGTGGCCACAACCAAGTGGCTCAGGCTCTGG 1560  
 AGAAGCTGCACTTCCCTCAGCAAGCTCAACATCTATGCAAT 1600  
 CAGGCTGGCAGGGCAAGCAAGGCAAGGCAAGGCTATGGCGGC 1640  
 TTGGCTCTGGGCCCCCGCAGGCTCTCAACCTGCTGCAGC 1680  
 TCTACAGCCATCTTTCTGCAACCTTGGCAAGGATGCCCCG 1720  
 AACTGGGTTTCTCAGGCTCCAGCAATCTTTGGCCACTACT 1760  
 CACAAGCTTCAAAACAGCAACAGCTACCTTGGCAATCAGC 1800  
 GCTTTCAAGGCTGTGCTGCTCAAGCTCTATGCTCT 1840  
 GCAAGCAATATAGGGGCTTACCTGGCCCCCTCAGCCGCT 1880  
 CGGTACAGTGGCCCCCTGCTGCTGCAAGCTTCCATCTCAA 1920  
 AACTTCCACTTCAAGCAAGGGCTCCGAGGCTACAGGCTAC 1960  
 CATGGCTGCAACAGGCAAGGCTTTTGGGCTATCTTTTGTAT 2000  
 ATGGAGTCAATTTTGTGTAATAAAGGCTGCAAGCTTAAAA 2040  
 AA 2080  
 AAAAAA 2087

FIG. 8B

## mmFATP3 protein sequence

AADPESESSEGC SLAWRLAYLAREQPTHTFLTHCAQRF SYAFAERESNRIA 50  
 RAFLRARGWIGRRGSGRGSTEEGARVAPPAGDAAARGITAPPLAPGATV 100  
 ALLLPAGPDEF LWIWFGLAKAGLRITAFVPTALRRGP L L HCLRSOGASALVL 150  
 ATEFTLESLEPDL PALRAMGH LWATGPEINVAGISNLLSEADQVDEFVP 200  
 GYLSAPQNMIDICLYIFTSGFTGLPKAARISHLKV LQQGFYHLCGVHQE 250  
 DVIYIALPLFYHMSGSLIGIVGCLGIGATVVLKPKFSASQFWDDQKHRVT 300  
 VFQYTIGELCRYLVNQPPSKAEFTLHKVRLAVGSGLRPD IWERFLRRFGFLQ 350  
 ILETYGMTEGNWATFN YTGRCQAVGRASWLYKHIFPFSLIRYIVMTGEPT 400  
 RNAQGHOMITSPGEFGLLVAFVSQQSPFLGYACAPELAKDKLLKDVFWSG 450  
 DVFFNTIGDLLVCD EQFLHEHDIRTGDTIRWKGENVATTEVAENVLEILDEL 500  
 QEVNTYGVTVFGHECRAGMAALALRPPQALNLVQLYSHVSENLPFYARPR 550  
 FLRLQESLATTEFTFKQOKVRMANEGFDPSVLSDFLYVLDQDTICAYLPLTP 600  
 ARYSALLSGDLRI 613

FIG. 9

## mmFATP4 DNA sequence

CCCCAGCGGTCCGCCCCAGCGGTCCGGCATCGGCCAAGCTGGG 40  
 CGTGGAGCGCGCTCTCATCAACACCAACCTTAGCGGGCAT 80  
 GCCCCTGCGCCACTGTCTTCACACCTCAAAAGGCACGAGCTC 120  
 TCATCTTTTGGCAGTCAGATCGGCTCAGCTATCTGTTCAGAT 160  
 CCATGCTAGCCCTGGACCCCACTCAGCCTCTTCTGCTCT 200  
 CCACTCCTGGCAGCCACAGCAGTCCCGCTCAGCAGACAGC 240  
 ATCTGCAACCTCTTCTGCAACATGCCCCGAAGCACTGCTG 280  
 CAGTCACCCAGCAAGGGTTTACAGATAAGCTCTTCTAC 320  
 ATCTACACATCGGGCAACACGGGGCTACCCAAAGCTGCCA 360  
 TTCTGGTGACACAGCAGGTTATATCGTATGGCTTCCCTGGT 400  
 GTACTATGCAATTCGGCATCGCGCTCATGACATTGTCTAT 440  
 CACTGCCCTCCCTCTCTACCACTCAAGCAGGAACATCGTC 480  
 GGCATTGGCAGTCTTACTCCAGGCATCACTGTGGTCAT 520  
 CCGCAACAGTTCTCAGCCTCCCGGTTCTGGCATCATTTGT 560  
 ATCAAGTACAACTGCACAGTGGTACAGTACATTGGCGAGC 600  
 TCTGCCCGCTACCTCTCTCAACAGCCACCCCGTCAGGCTCA 640  
 GTCTCGGCACAAAGGTGCCCATGGCACTGGCCAAAGGCTCT 680  
 CGGCAGTCCATCTGCAACCACTTCTCCAGCCGTTTCCACA 720

FIG. 10A

TCCCCCAGGIGGCTCAGTTCTATCGGGGCACTCAATGCAA 760  
 CTGTAGCCCTGGGCAACTTTTCACAGCCGGGIGGGGGCTGT 800  
 GGCCTCAATAGCCGCATCCTGTCCCTTTGTGTACCCCTATCC 840  
 GTTTGGTACGTGTCAATCAGCATACCATGCACACTCATCCG 880  
 GGCACCCCATGCGAGTCTGCCATTCCCTGTCAACCAGGTACG 920  
 CCAGGCCAGCTGGTGGGTGGCATCATCCAGCAGCACCCTC 960  
 TCGGCCCGTTTCGACGGGTACCTCAACCAGGGTGGCAACAA 1000  
 CAACAACATTGCTAATGATGTCTTCAACAGGGGGCACTAA 1040  
 GCGTACCTCACTGGTCAAGTCCCTGGTCAATGAGCTGG 1080  
 GTTACCTGTACTTCCGACATGGCACTGGGGCACAGTTCCG 1120  
 CTGCAAGGGGACAAATGTATCTACCACTCAGGTGGAGGCG 1160  
 ACACCTCAGCCGCTGCTTCATATGGCAGATGTGGCAGTTT 1200  
 ATGGTGTTCAGGTGGCAGCAACTCAAGCCCGAGCAGCAT 1240  
 GGCCTGCCGTTGCAAGTCCCATCAGCAACTGTGCACTCGAG 1280  
 AGCTTTGCAACAGACCTTCAAAAAGGAGCTGCCCTCTGTATG 1320  
 CCGCGCCCATCTTCCCTGGGCTTCTTGGCTCAGCTGCACAA 1360  
 CACAGGGCACTTCAAGTTCACACACAGAGTTGGCGAG 1400  
 CAGGGCTTTCACCCATCTGTGTGTCAAGCAACCCGCTGTCT 1440  
 ATCTGGATGCTCGCAAGGGCTGCTACGTTTGCCTGCAACA 1480  
 GCAAGCCCTATACCCGCATCCAGGCAGCCAGCAAGCCTG 1520  
 TCAATTTCCCCCTACATCCCTCTCAGGGCCACACATGCTG 1560  
 CATTCAGAGCCCTAGCGTCCACCCCAAGGGTCCCTGGCA 1600  
 ATGCCAGACCAAGCTAGCAGGGCCCGCAACCTCCGCCCCCT 1640  
 AGGTGCTCATCTCCCCCTCTCCCAAACTGCAAGTCACTCA 1680  
 CTGCGCGCTTCCCCCAACCTCCCAAGGCTTCTGTCAAGT 1720  
 CTCTATCCAAAGCTGTGTCTTCTGGTCCAGGGCTGGCCCCCTG 1760  
 GCCCCAGGGCTTCTCATAGGCTCCCTTTAGCTTGGTATCTT 1800  
 GCGTCCACCGGGCTAGGGTGTGGCAAGCACTCACTAACA 1840  
 TCCCTCCAAATCACAAGCGAGCTTACAAAGCAACCAAGCCA 1880  
 AAGCCTGTACACTCAGCAAGCTAAGTGGCCAGCACTATA 1920  
 GTGGCCAGTCAATCCCATGTCCACAGAGATCTTTGGTCCAG 1960  
 AGCTGCCAAAGTGTCACTCTCTCCCTGGCTGCACTCTGGG 2000  
 CAAGAAGAGCAAGCATGTGTGGCCAGTGGGCACCTGTCTCAA 2040  
 CAAGTCAGCATCACAACCTCAGTCCCTTGTCTTCTCCAGTT 2080  
 CCGTGTCTCTGTCTCTGGGGCAGGCAGCCAGCTGTCTCTG 2120  
 TCTGTCTTCTCTGGCTGTCTGTGAGTCTGTGTGTCTCTCT 2160  
 CATCTGTCTTACGCTCAGTGTGGGTGCAACAGGCATCAGG 2200  
 ACAGTGTGGCTCAGGGGCCAATAAAGCTCTGCTTGAATCC 2240  
 TCTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2280  
 AAAAAAAAAAAAAAAAAAAAAA 2301

FIG. 10B

## mmFATP4 protein sequence

HASAHASGMAKLGVEAALININLRRDALRHCLDTSKARAL 40  
 IEGSEMASAICETHASLEPTLSLFCSGSWEPSTVPVSTIEH 80  
 LDPLILEDAPKHLPSHPDKGFTDKLFYIYISGTTGLPKAAT 120  
 VVHSFYRMASLVYYGFRMRREDDIVYDCLPLYHSSRKHRG 160  
 DWQCILLHGMIVVIRKKFSASRFWDDCIKYNCTIVVQYIGEL 200  
 CFVLLNQPPREAESRHKVRMALGNGLRQSTIWIDFSSREHI 240  
 PQVAEFYGCATEONCSLGNFDSRVGACGFNSRILSFVYPIR 280  
 LVRVNEDIMELIRGPDGVCTPCQPCQPGQLVGRITIQQDPL 320  
 RRFDDGYLNQGANNKTIANDVFKKGDQAYLTGDLVLMDELG 360  
 YLYFRDRICGTFRWKGENVSTIEVEGILSRLLHMAIVAVY 400  
 GVEVPGIEGRAGMAAVASPISNCLLESFAQTLLKKEPLVA 440  
 RPIFLRFLPELHKITGTFKFQRTIELRKEGFDPSVWKDPLFY 480  
 LDARKGCYVALDQFAYTRIQAGEEKL 507

FIG. 11

## mmFATP5 DNA sequence

CACTCATCAGAGCTAAGACAGACTACACGCTCTCATCTAC 40  
 TTTCAGAAAGAGCCCAATGCCATGGGTAATTTGCAAGAACTA 80  
 ACCTTACTGCTGTGTGCTGCTTCTGCTGCTTGGCCCTGGGGC 120  
 AGCCCCCATGCGCCAGCAGCTATGCGCTCTGGCCCTGGCGTTC 160  
 GTTCCCTGGGCAACCCCATATGCCCTTGTGCTGCTTGGCTTTC 200  
 GCATTGCTGCGGCAACCCCTGGATCAGCTCCTGCAATGCCCC 240  
 ACTGCGCTCAGCCCTGGTACCAAGCAGCTCTTACCTTATTCCT 280  
 ATTGCCCTCTACAGCCACCCCCAGGGCTACGCTGGCTGGCAT 320  
 AAAGATGTGGCTTTTACCTTTCAAGATGCTTTTCTATGGCC 360  
 TAAAGTTCAGGCGACGCCCTTAAACAAACATCCTCCACAGAC 400  
 CTTTGTGCTATGCTTTAGAGCGGCAAGCACTGGCATGGGCT 440  
 CACCGGGTGGCCCTGGTGTGTACCTGGGCTCTCAGGGCTCCT 480  
 CAATCACAATAAGCCAGCTGCAATGCCAGGTCTGTGACAGCC 520  
 AGCATGGCTCCTGAAAGCAAGCGCAAGCATGCCGTAATC 560  
 CAGAACACAAGCAATGCTGCTGCTATCTTATGTTCTCCCGT 600  
 CCAAGCAACATTTCCTTTCAGTGTGCTTTCAGGGCTTGGC 640  
 CAAGTTGGGCTGGCCCTGTGGCTGGATCAATCCACACAGC 680  
 CGAGCGATGCCCTTGCCTACACTGTGTACGCAAGCTCTGGCG 720  
 CCAGTGTGCTGATTGTGCTATCCACACCTCCAGTACAACT 760  
 GCAACAGTCCCTTCCCAAGCTGCTAGCTGCAACATTTCAC 800

FIG. 12A

TGGTTCTACCTTTGGGACACAGCTCAACCTACCCCGGCGAGTAG 840  
 AGGCTCTGGCGAGCTTCCCTGGATGCTGCACTTCTGCAACC 880  
 AGTACCTTGGCAGCCTTCCAGCTACGATTAAAGTGGAAATCT 920  
 CCTGGCATATTTCATCTTTACCTTCAGGCAACCACTGCACTCC 960  
 CAAAGCCAGCCATCTTTATCACTGACCGGGTCATACAAGT 1000  
 CAGCAACCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1040  
 CTGGTCTATCAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1080  
 TTCT 1120  
 CT 1160  
 GCTCAAGTGGCGGCGAGCATGCGCTAACAGTCTCTCTCTCTCT 1200  
 TGGGTCAAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1240  
 ACCACAAACAAACATACTACAGTCTCTCTCTCTCTCTCTCTCT 1280  
 ACTGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1320  
 GCTTTGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1360  
 ACAGGGCAATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1400  
 GGGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1440  
 CT 1480  
 GCT 1520  
 CCAGCAAAAGCCAGCACTCTCTCTCTCTCTCTCTCTCTCTCT 1560  
 ACCAACCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1600  
 CAATGGCAAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1640  
 CT 1680  
 AAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1720  
 CCGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1760  
 TCT 1800  
 TCTATGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1840  
 CATGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1880  
 CCGCAAGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1920  
 CCTATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1960  
 CCACT 2000  
 CT 2040  
 TCT 2080  
 CATCT 2120  
 AATCT 2160  
 TCT 2200  
 CT 2240  
 CT 2277

FIG. 12B

22/202  
mmFATP5 protein sequence

MALALRWFLGDPTCLVLLGLALLGREWLSSWMPHWSLVG 40  
AALILFLPLPLQPPFGLRWLHKDVAFTEKMLFYGLKFRRL 80  
NKHPPPEIFVDALERQALAWPDRVALVCTGSEGSSTINSQL 120  
DARSCQAAWLKAKLKDAVIQNIQDAAAILVLPSTISAL 160  
SVFLGLAKLGCPVAWINPHSRGMPLIHSVRSSCASVLVD 200  
PDLOENLEEVLEKLI AENIHCFYLGHSSPTGVEALCASL 240  
DAAPSDFPASLRATIKWKSPAIFITFTSGTIGLPKPATLS 280  
HERVIQVSNVLSFCGCRADDVVDVLPVYHTTIGLVGLG 320  
CLQMGATCVLAPKFSASRFWAECRQHGVIIVLYVGEILRY 360  
LCNVPEQPEDKIHTVRLAMGTGLRANWKNFQORFGPIRI 400  
WEFYGSTEGNGLMNYVGHGCAVERTSCILRMLTPPELVQ 440  
FDIETAEPLFDKQCFCTFVEFCKPGLLLIKVRKNQPFLEY 480  
RGSQAESNRKLVANVRVGLDYFNTGDLVILDDQEGFFYFQ 520  
DRLGDTFRWKGENVSTGEVECVLSSLDLEEMNVYGVFVP 560  
GCECKVGMMAVKLAPCKTFDQKLYQHVRSWLPAYATEHF 600  
IRIQDSLETINTYKLVKSRLVREGFDVGTIADPLYTILNK 640  
AQTFERSLMFDVYQAVCEGIWNL 663

FIG. 13

hsFATP2 DNA sequence

ATGGCATTGACTCTTTTCCCTGGACAAAGTGGATCAAGTATC 40  
AACTCAACCTATCCCAAGTATGCGAGGCTCTCAAGTCACT 80  
TTTTCCACTCCCTGCTTATACATTTATACCTTCCTGGAACCA 120  
CAGGTCCTTCCAAAAGCAGCCATGATCACTCATCAGCCCAT 160  
ATGGTATGCAACTGCGCTCACTTTTGTAGCGGATTGAAG 200  
GCACATCATGTATCTATATCACTCTGCGCCCTTTTACCA 240  
GTGCTGCACCTACTGATTGGCATTCAAGGATGTTATGTTGGC 280  
TGGTGCCTACTCTTGGCTTGGGCACTAAATTTTCAGCCAGC 320  
CAGTTTGGGATGACTGCGCAAAATACAAAGTCACTGTCA 360  
TTCAGTATATCGGTCAACTGCTTGGGTATTATGCAACTC 400  
ACCAACAGAAACCAATGACCGGATCATATAAGTCACTG 440  
GCACTGGCAATGCGTTACAGGAGATGCTGCGCAACAAT 480  
TTGTCAAGCATTTCGGGCATATGCTATGAGTTCTA 520  
TGCCTGCCACGCAAGCCATATTGCAATTTATCAATTATGCG 560  
ACAAAAGTTGGTGGCTGTGCAAGAGTAAACTACCTACACA 600  
AAAAAATCATAACTTATGACCTGATTAATATGATGTGCA 640  
CAAACATCAACCTGTCCGTCATCAAAATGCATATTGCGTC 680  
ACAGTTCCCAAGGTCAGTTGCACTTCTGGTTTGCAGAAA 720  
TCACACAACITACACCATTTAATGGCTATGCTGCGAGAAA 760  
GGCTCAGACACAGCAAGAAAACCTCAGCATGCTCTTTAAG 800

FIG. 14A

AAAGGAGACCCCTCTATTTCAACAGTGGAGATCTCTTAATGG 840  
 TTGAOCATGAAATTTTCATCTATTTCCACACAGTTGG 880  
 ACATACATTCGGGTGGAAAGGGGAAAATGTTGGCCACCACT 920  
 CAAGTTGCTGATATAGTTGGACIGGTTTCATTTTTTTTCCAA 960  
 GGAAGTAAATGTTTATGGCAGTGCATGGGCCAAGATNAT 1000  
 GCAGGTTTCAATTTGGCATGGGNTTCCNTTCAAAATGCAAA 1040  
 CAAAACCATGCAATTTCATGCAACAAATTTTTTTCAGNAC 1080  
 ATTTGCTGATAACCNACCTAGTTATGCAAGGCCCCGGTTTT 1120  
 NTAAGAANACAGGACACCATTTGCAATCACIGCAATTTTTTA 1160  
 AACACCGCAAAATGACCTTTGGTGGAGGAGGGCTTTTAACC 1200  
 CNGCIGTCATCAAAAGATGGCTTGTATTTTCTTGGATGACA 1240  
 CAGCAAAAATGTATGTGGCTATGACTGAGGACATNTATAA 1280  
 TGGCATAAGTGTAAACCCCTGCAATTTNTCAATATTTCCCA 1320  
 GGAGGATAATTCAACATTTCCACAAAGCAACIGCAATGGAC 1360  
 AGCCACTTGATATAATCCAACTTTAATTTGATTGCAAGATT 1400  
 GTCAGCAAAATTTTGTAGCAATTTTGCATACCCGTAAAGGG 1440  
 AGACTTTTTTAAATAACAGTTGAGTCTTTTGCAGTAAAAA 1480  
 CATTTAGACATTATTATTTTTTCAGTGTGCACTACIGTTT 1520  
 GTATTTCGAAACTGAGCTTTGTTGAGGGGAGGCATTATTT 1560  
 TTTAAATACTTAGTAAATTAAGCAACACCAACATGTCAA 1600  
 AAAAAAAAAAAAAAAAAAAAAA 1622

FIG. 14

## hsFATP2 protein sequence

YIYTSGITGLPKAAMITHQRIWYGTELTFVSGLKADDVIY 40  
 IITLFFYHSAALLIGTHGCIIVAGATIALRIKFSASQFWDCC 80  
 RKYNVIVIQYIGELLRYLONSPQKPNDRLHKVRLALGNEL 120  
 RGDWNRQFVKRFCDICTYEFYAATEGNIGFMNYARKVGAV 160  
 GRVNYLQKKLITYDLIKYDVEKDEFVRDENGVCVRVPKGE 200  
 VGLLVCKTITQLTPFNGYACAKAQTEKKLRDVEFKGLDYF 240  
 NSCDLLMVDHENFTLYFHLRVGDIFFWKGENVATTEVADIV 280  
 GLNDEF 286

FIG. 15

## hsFATP3 LNA sequence

CAATTGGGCAACCCCAAGGGGCACTGTATGGCCATCTCC 40  
 AGGTGAGCCAGGGCAAGTTGCTAAAGCAATGTCCTTCCGCC 80  
 TGGGATGTTTTCTTCAACACTGGGCACTGCTGCTGCTGCTG 120  
 CATCAACCAAGTTTTCTCCGCTTCCATCATCTGCTGCTGCTG 160

FIG. 16



## mmFATP5 protein sequence

MAIALRWFLGDPTCLVLLGLALLGRFWISSWMEHWSLVG 40  
 AALTLFLLLPLOPPGLRWLHKDVAFTEKMLFYGLKERRRL 80  
 NKHPPEITFVDALERQALAWPDRVALVCTGSEGSSTINSOL 120  
 DARSCQAAWVLKAKLKDAVIQNIRDAAAIIVLPSKTTISAL 160  
 SVFLGLAKLGCPVAWINEHSGMPLIHSVRSSCASVLTVD 200  
 PDIQENLEEVLKLLAENIHCFYLGHSSTPFGVEALCASL 240  
 DAAPSDPVPASLRATIKWKSPAIFITFTSGTTGLEPKPATLS 280  
 HERVIQVSNVLSFCGCRADDVVDVLPYHTTIGLVLGFTLG 320  
 CLQMGATCVLAPKFSASRFWAECRQHGVIIVILYVGETLRY 360  
 LCNVPEQPPELKIHVRLAMGTGLRANWKNFQORFGPIRI 400  
 WEFYGSTEGWGLMNYVGHGCAVGTSCILRMLTIPFELVQ 440  
 FDIETAEPLRQKQFCIFVEFGKPGILLIKVRKNQPFLLGY 480  
 RGSQAESNRKLVANVRVGLDYFNTGDLTLDQEGFTYFQ 520  
 DRLGDTFRWKCHNVSTGEVECVLSSIDFLEENNVYGVFVP 560  
 GCECKVGMMAVKLAPGKTTFDGOKLYQHVRSWLPAYATPHF 600  
 TRIQDSLEITINTYKLVKSRLVREGFDVGLTADPLYTLINK 640  
 AQIFRSIMPLVYQAVCEGIWNL 663

FIG. 13

## hsFATP2 DNA sequence

ATGGCATTGACTCTTTTCCCTGGACAAAGTGGATCAAGTATC 40  
 AACTCAACCTATCCACAGTTCATGCGAGGTCTCAAGTCACT 80  
 TTTTCCACTCCCTGCTTATACATTTATACCTCTGCAACCA 120  
 CAGGTCTTCCAAAAGCAGCCATGATCACTCATCAGGGCAT 160  
 ATGGTATGCAACTGGGCTCACTTTTGTAAAGCGGATTCAAG 200  
 GCAGATGATGTCATCTATATCACTCTGCCCCCTTTTACCACA 240  
 GTCCTGCACTACTGATTGGCATTCAAGGATGTATTGTGGC 280  
 TGGTCTACTCTTGGCTTGGGACCTAAATTTTCAGCCAGC 320  
 CAGTTTTCGGATGACTGCGACAAAATACAAAGTCACTGTCA 360  
 TTCAGTATATCGGGTCAACTGCTTGGGTATTTATGCAACTC 400  
 ACCACAGAAACCAAAATCAACCGTGCATCAAAAGTCAACTG 440  
 GCACTGGCAAAATGGCTTACCAAGGAGATGTGTGGCAACAAT 480  
 TTGTCAACACATTGGGGCATATGTCATCTATCACTTCTA 520  
 TGCTGCCACCTGAAGGCAATATTGGCATTATCAATTATGCG 560  
 ACAAAAGTTGGTCTGTTCGAACAGTAAACTACCTACACA 600  
 AAAAAATCATAACTTATCACTGATTAAATATCATGTGCA 640  
 CAACATCAACCTGTCCGTGATCAAAATGCATATTGGGTC 680  
 ACAGTTCCCAAGGTCAAGTTGCACTTCIGGTTTGCAGAAA 720  
 TCACACAACCTACACCATTTAATGGCTATGCTGCAAGAAA 760  
 GGCTCACAACACACACAAAACCTCAACATGTCTTTAAG 800

FIG. 14A

AAAGGAGACCTCTATTTTCAACAGTGGACATCTCTTAATGG 840  
 TTGAACCATGAAAATTTTCATCTATTTCCACGACAGAGTTGG 880  
 ACATACATTCOOGGIGCAAAAGGGCAAAATGTGGCCACCACT 920  
 CAAGTTGCTCATATAGTTTGCACCTGGTTGATTTTTTTTCCAA 960  
 GCAAGTAAATGTTTTATCGGAGTGCATGGGCCAAGATNAT 1000  
 CGAGGTTTCAATTGGCATGGGNTTCNTTCAAAATGCAAA 1040  
 CAAAACCATGCAATTTCATGCAAGCAAAATTTTTTTCAGNAC 1080  
 ATTGCTCTATAACCNACCTAGTTATGCAAGGCCCCGGTTTT 1120  
 NTAAGTAANACAGGACACCATTCAGATCACCTGCAATTTTTTA 1160  
 AACACCGCAAAATGACCTTTTGGTGGAGGAGGGCTTTAACC 1200  
 CNGCTGTTCATCAAAAGATGCCCTTGTATTTTCTTGGATGACA 1240  
 CAGCAAAAATGTATGTGGCTATGACTGAGGACATNATATAA 1280  
 TGCCATAAGTGTATAAAACCTGAAATNTTGAATATTCOCA 1320  
 CGAGGATAATTCAACATTTTCCAGAAAGCAAACTGAATGGAC 1360  
 AGCCACTTCTATATAATCCAACTTTAATTTTGAATGAAGATT 1400  
 GTGAGCAAAATTTTGTAGGCAATTTGCATACCCGTAAAGGG 1440  
 AGACTTTTTTTAAATAACAGTTGAGTCTTTTGGCAAGTAAAAA 1480  
 CATTTAGACATTATTATTTTTTCACTGTGCACTTACTGTTT 1520  
 GTATTTCGAAACTGAGCTTGTGTGGAGGCAAGGCATTATTT 1560  
 TTTAAATAACTTAGTAAATTAAGCAACACCAACATGTGAA 1600  
 AAAAAAAAAAAAAAAAAAAAAA 1622

FIG. 14B

## hsFATP2 protein sequence

YIYTSGITGLPKAAMTHQRIWYGTGLTFVSGIKADDVIY 40  
 ITLPFYHSAALLIGIHGCIVACATLALRIKFSASQFWDCC 80  
 RKYNVIVITQYIGELLRYLONSPQKENDRDHKVRLALGNEL 120  
 RGDWWRQFVKRFEDICTYEFYAATEGNIGFMNYARKVGAV 160  
 GRVNYLQKKLITYDLIKVDVEKDEFPVRDENGVCVRVPKGE 200  
 VGLLMCKTIDQLTPFNGYACAKAQTEKKKLRLDVEKKEDLYF 240  
 NSGDLMLVHENFTYFHLRVGDIFFRWKGENVATEVADIV 280  
 GLVDEF 286

FIG. 15

## hsFATP3 DNA sequence

CAATTGGGCAACCCCAAGGGCACTGTATGGCCATCTCC 40  
 AGGTGAGCCAGGGCAAGTTGCTAAAGCATGTCTTCCGGCC 80  
 TGGGATGTTTTCTTCAACACAGGGCACTGTCTGTCTGC 120  
 CATCAACCAAGGTTTTCTCCGCTTCCATCATCTGCTGAG 160

FIG. 16A

ACACCTTCAGGTGCAAGGGGAGAATGTGGCCACAACCGA 200  
 GGTGGCAGAGGTCTTCCAGGCGCTACATTTTCTTCAGGAG 240  
 GTCAACGCTCTATGCAATCAGTGTGGCCAGGGCATCAAGGCA 280  
 GGGCTGCAATGGCAGCGGCTAGTTCGTGGGTCCCCCCCCACGC 320  
 TTTGGACCTTATGCAGCTCTACACCCAGGTGTCTGACAAAC 360  
 TTGGCACCTTATGCCCCGGCCCCGATTTCCTCAGGCTCCAGG 400  
 AGTCTTTTGGCCACCACAGACACCCCTTCAAAACAGCAGAAAGT 440  
 TCGCATGGCCAAATGAGGGGCTTCGACCCCAAGCACCCTGTCT 480  
 CACCCACTGTACGTTCTGCAACAGGCTGTAGGTGGCTTACC 520  
 TGCCCCCTCACAACCTGCCCCGGTACAGCGGCTTCTGGCAGG 560  
 AAACCTTTCGAATCTGCAACCTTCCACACCTGAGGCACTTG 600  
 ACACAGCAACTCTGTGTGGGGTGGGGGGGGGTTCAGGTGTAC 640  
 TGGGCTGTACGGCATCTTTTCTATACCACTGCGGCTCA 680  
 CTATTTTGTAAATAAATGTGGCTGCACTGATCCAGCTGTCT 720  
 TCTGACCTTACAAAAAATAAAAAAATAAAAAA 753

FIG. 16B

## hsFATP3 protein sequence

QFGTPRGIVWFHLQVSQKLLKIDVERPQVFFNTGDLIVC 40  
 DDQGFLEFHLRTGDLFRWKGENVATTEVAEVFFALDFLQE 80  
 VNVYGVIVPGHEGRACMAALVLRPFHALDLMQLYTHVSEN 120  
 LPPYARPRFLRLOESLATTETFRKQKVRMANEGSDPSTLS 160  
 DPLVLDQAVGVLPITTRYSALLAGNLRT 191

FIG. 17

## hsFATP4 DNA sequence

TCAAGTACAACCTGCACCAATTGTGATANCATTGGTGAACTG 40  
 TGCCGNTACCTTCTGCAACCAGCCACCCGGGGAGGCACTAA 80  
 ACCAGCAACCAGGTTCGGCATGGCACTAGGCAATGGCTTCCG 120  
 GCAGTCCATCTGCAACCAACTTTTCCAGCGGCTTCCACATA 160  
 CCCCAGGTGGGCTCAGTTTATCGGGGGCCACAGAGTGCAACT 200  
 GTAGCCTTGGGCAACTTCCACAGCCAGGTGGGGGCTGTGG 240  
 TTTCAATAGCGGCATCCTGTCTTCTGTGTACCCCATCCGG 280  
 TTGGTACGGTGTCAACCAAGCAACCATGCAAGCTGATCCGG 320  
 GGGCGCAAGGGGCTGTGCATTCCTTGGCAGCCAGGTGAGCC 360  
 GGGCCAGCTGGTGGGGCGCATCATCCAGAAAGACCCCTTG 400  
 CGCGGCTTCCATGGCTACCTCAACCAGGGGGCAACAACA 440  
 ACAAGATTGGCAAGCATGTCTTCAAGCAAGGGGCAACAGGC 480  
 CTACCTTACTGGTCAATGTGCTGGTCAATGCAACCTGGGC 520

FIG. 18A

TACCTGTACTTCCGAGACCGCACTGGGACACGTTCCGCT 560  
 GCAAGGTCACACCGTGTCCACCACCGAGGTGCAAGGCAC 600  
 ACTCAGCCCGCTGCTGCAATGGCTGACGTGCGCGTGTAT 640  
 GGTGTCCAGGTGCCAGCAACCGAGGGCCGGGGCCGCAATCG 680  
 CTGCTGTGGCCAGCCCCACTGGCAACGTGTACCTTGGCAGC 720  
 GCTTTGCTCAGGTC 734

FIG. 18B

## hsFATP4 protein sequence

IGELCRYLLNQPPREAFNQHQVRMALGNGLRQSIWINFSS 40  
 RFHTLPQVAEFYGAIEONCSLGNFDSQVCAAGNSRILSFV 80  
 YPIRLVVRVNEIDIMELIRGPDGVCIPQCPGEPGQLVGRLLQ 120  
 KDPLRRFDGYLNQGANNKIADVFKKEDQAYLTGDLVM 160  
 DELGYLYFRDRTGDIFFRWKGENVSTTEVEGTL SRLILIMAD 200  
 VAVYGVEVPGIEG 213

FIG. 19

## hsFATP5 DNA sequence

CNIGCCCTCTTTGTACCAAGTCATGGCACTTTGTGCTTGGCA 40  
 TTCTCCGGCTGCTTACATCTCCGACGACCTGTGTCTGCC 80  
 CCCCAGTTCTCTACTTCCCTGCTTCTGGCATCACTGTCCG 120  
 CAGCATGGCGTGACAGTGATCTCTGTATGTGGCGAGCTCC 160  
 TGCCTACTTGTGTACATTTCCCCAGCAACCAAGCAACCG 200  
 CACACATACAGTCCCGCTGGCAATGGGCAATGCACTACCG 240  
 GCTCATGTGTGTGGGACACCTTCCAGCAGCGTTTGGTCTT 280  
 ATTTCCGATCTINGGCAAGTCTTACGGGCTTCCACACAAGG 320  
 GCAACATGGGGCTTTAGTTCAACTATTGTGTGGGGCGGCTG 360  
 CGGGGSCCTGCGGGCAAGATGCAAGCTTGCCTCTCCGAA 400  
 TGGTGTCCCCCTTTTACGCTGGTGCACTTCCATGCAAGGC 440  
 GGCGGAGCCTGTGACGGCAATCAGGGCTTCTGCATCCCT 480  
 GTAGGGCTACGGGAGCGGGGGCTGCTGTTCACCAAGGTGG 520  
 TAAGCCAGCAACCCCTTGGTGGGCTACCGCGGCCCCCGACA 560  
 GCCTGTCCGAACCGAAGCTGGTGGCAACGTGGGGCAATCG 600  
 GGCGAAGCTTACTACAAACACCGGGCAAGTACTGGCCATGG 640  
 ACCCGCAAGCCCTTCTCTACTTCCGCAACCACTGGGGCA 680  
 CACCTTCCGATGCAAGGGCGACAACGTGTCCACGCACGAG 720  
 GTGGACGGGGGTGTGTGTGGCAGGTGCACTTCTTGCACAGG 760  
 TTAAAGTGTATGGCGTGTGGTGGCAGGTGTGTACGGGTAA 800  
 GGTGGCCATGGCTGCTGTGGCATTAGCCCCCGGCGCACT 840

FIG. 20A

TTGGAGGGGAGAGGTTGTACAGCAAGTTGGGGCTTGGC 880  
 TCCCTGCCCTACGCTACCCCCCATTTTCATCCGCATCCAGCA 920  
 CGCCATGCGAGGTACACAGCAGGTTCAAACCTGATCAGACC 960  
 CGGTTGGTGGCTGAGGGCTTCAATGTGGGGATCGTGGTTG 1000  
 ACCCTCTGTGTTGTACTGGACAACCGGGGGCAGTCCCTTCGG 1040  
 GCGGCTGACGGCAGAAATGTACAGGCTGTGTGTGAGGCA 1080  
 AACTGGAGGCTCTGATCAGCTGGCCAAACCACTGGGGTAG 1120  
 GGATCAAAGCCAGCCACCCCAACCAACACACTGGGTGT 1160  
 CCGTTTCATCCTGGGGCTGTGTGAATCCAGCCTGGCCAT 1200  
 ACCCTCAACCTCAGTGGGCTGTGAATGACAGTGGGGCTG 1240  
 TAGCAGTGGCAGATAAACTCAGMTGYGTTCACAGAA 1278

FIG. 20B

hsFATP5 protein sequence

EGQHGALVQLLIGALRGFGCKDCAQLLRMLSPFELVQFLM 40  
 EAAEFVRINQCFCTPVELGEPELILLIKVWSQPFVGYRCP 80  
 RELSERKLVNRNVRQSEVYYNIGLVNLAMREGEFLYFRDL 120  
 GDIFRWKGENVSTHEVEGVLSQVDFLQQVNVYGVCFEGCE 160  
 CKVGMAAVALAPGQTFDGEKLYQHVRWLPAVATPHFIR 199

FIG. 21

hsFATP6 DNA sequence

CGCTTGTGTGTGTAAGCAAGCAATTTTTCAGCAAGCCAGTTT 40  
 TGCAGTGCACGTGCAAGCAAGTATCATGTGACTGTGTTTCAGT 80  
 ATATTGGCAACTTTGTGGCTAACCTTTCGAAACAATCTAA 120  
 GACACAGGCAAAAAGCATCATAGGCTGGCTTTGGCAATT 160  
 GCAAAATGGCATAACGAGTCTATGATGCACAGATTTTTATG 200  
 ACAGATTTTGGCAAAATATAAAGGCTGTGTGCAACTTTATGCAGC 240  
 TACCGAATCAAGCATATCTTTTCATCAACTACACTGGCACA 280  
 ATTGGAGCAATTGGCAGCAAAATTTGTTTTACAAACTTC 320  
 TTTCACCTTTTCACTTAATAAAGTATCACTTTTCAGAAACA 360  
 TCAACCCATGACAAATCAGCAGGGTGGGTATTCATGACA 400  
 AAAAGGCAACCTGGACTTCTCATTTCTCGAGTCAATGCCAA 440  
 AAAATCCCTTCTTTGGCTATGCTGGGGCTTATAAGCACAC 480  
 AAAAGCAAAATTCCTTTTGTCTATGTTTTTAAGCAGGCAGT 520  
 GTTTAACCCTAATACTGGCACTTAATAGTCCAGGATCAGG 560  
 ACAATTTCCCTTATTTTTTGGCAACCTACTGGCAGCACTTT 600  
 CAGATGCAAAAGCAGAAATGTGGCAACCACTCAGGTTGCT 640  
 CATGTTATTCGATGTGTGCTTTTCATACAGCAAGCAAAAG 680  
 TCTATCGGTGGCTATATCAGCTTATCAAGCAAGCAGG 720

FIG. 22A

AATGGCITTCIATTATTTTAAACCAATACATCITTTAGAT 760  
 TTGGA AAAAGITTTATGAACAAGITGTAACATTTCIACCAG 800  
 CITATGCTTGICCAAGATTTTAAAGAAATTCAGCAAAAAT 840  
 GCAAGCAACAGCAACATTCAAACTATTGCAAGCATCAGTTG 880  
 GTGGAAGATGGATTTAATCCACTGAAAATTTCTGAACCAC 920  
 TTTACTTCATGCATAACITTCAAAAAGTCTTATGTTCTACT 960  
 CACCAGGCAACTTTTATGATCAATAATGTTAGGGCAATA 1000  
 AAACITTTAAGATTTTATATCTAGCAACTTTCATATGCITT 1040  
 CITAGCAACAGTCAAGAGGGGGGATATGATTTCITTTATCAA 1080  
 ATGGGCAAGGGAGCTAACATTAATTATGCATGTAATA 1120  
 TTTCCCTTAATATGACACATAATTTTTTTAATTGCATAAGAA 1160  
 TTTTAATTTCTTTTAATTGATATAAACAGAGITGATTATT 1200  
 CTTTTTATCTATTTCGAGATTTCAGTGCATAACTAAGTATT 1240  
 TTCCCTTAATACTAAAGATTTTAAATAATAAATAGTGGCTA 1280  
 GCGGTTTGGACAATCACTAAAAATGTACTTTCTAATAAGT 1320  
 AAAATTTCTAATTTTCAATAAAAGATTAAATTTTACTGAA 1360  
 A 1361

FIG. 22B

## hsFATP6 protein sequence

ACVLKKKFSASQFWSDCCKYDVIVFQYIGELCRYLCKQSKREIEKIDHKVR 50  
 LAICNGIRSDWREFLDRFCNIKVCELYAATESISFMNYTGRICAGRT 100  
 NLFYKLLSTFDLJKYDFQKDEMRNEQGWEMRKRRPGLLSRVNAKNPF 150  
 FGYAGPYKPKDKLLCDVFKKGVYLNIGDLIVQDQINFLYFWLRTGDTF 200  
 RWKGENVATTEVADVIGMLDFIQEANVYGVAISGYEGRAGMASTILKENT 250  
 SLDLKQVYEQVVFPLPAYACPRFLRIQEKMEATGIEFKLLKHQLVEDGPNP 300  
 LKISEPLVFMDNLKKSIVLLIRELYDQIMLGELKL 335

FIG. 23

## mtFATP DNA sequence

TAGTCCATAACGTCAGGACCGCTCTGCGGGCCTGCGCACC 40  
 TTCCCTCAGCTTGGTCCACAAGCAATTCCACATTTCCCTAAA 80  
 CGAATCGACGGCTTACGTGTCCGATTACTACGGGGGGCCCA 120  
 CACACAACGGTCAAGGCTCATCGACCTGGCAACTCGGATGC 160  
 CGCGACGTGTTGGCGCACACCGCGGTCAATTGTGGGTGGGGC 200  
 AATCACCGGGCTGCTGGCGCGCGCGCAATTCCAGGGGTCC 240  
 ATCGGCACGGTGTTCAGGACCGGGCGCGGTCCGCTACGGTTC 280  
 ACGGACGTCTTCCCTCAATTCCGGCATCAGCAGCTCACTA 320  
 CCGGACGGCTTACCGCTACCGGCAACCGGTACGGCGGGTTC 360

FIG. 24A

FIG. 24B

## mtFATP protein sequence

```

msdyyyggahttvrlidlatmprvladtgvivrgamtgll 40
arpnskasigtvfgdraarygdrrvflkfgdqqltyrdana 80
tanryaavlaargvgpgdvvgimlmspstvlamlatvkc 120
gaiagmlnyhgrgevlahslglldakvliaesdlvsavae 160
cgasrgrvagdvltveäverfättapatmpasasavcaki 200
tafyiftsgttgfpkasvmthhrwlravfggmglrlkg 240
sdtlyscplplyinnaltvavssvinsgatlalgksfsasr 280
fwäevianratafvyigeicryllnqpakptdrahqvrvi 320
cgnglrpeiwdettrfgvarvcefyaaasegnsafinifn 360
vprttagvsmplafveyildtgöplrdasgrvrrvpdges 400
glllsrvnrllqpfdgytdpvasekklvinafrögdowfnt 440
gövmsspqgmghaafvdrlgdtfrwkgenvattgveaalas 480
dgtveectvygvqiprtgggragmaaitlragaeafdgaala 520
rtvyghlpgyalplfvrvvgsлахtttfksrkvelmqay 560
gadileöplyvlagpdögyvpyyaeypeevslgrrpog 597

```

FIG. 25



1	tcg acc cac ggc gtc cgg gac ccc aaa gca gaa gcc cgc aca gta ggc aca gcg cac cca	
61	aga agg gtc cag gag tct gca gaa aca gaa agg tcc ccg gcc tca gcc tcc tag tcc ctg	
121	cct gcc tcc tgc ctg agc ttc tgg gag act gaa ggc acg gct tgc agc ttc agg atg cgg	M R
181	gct ccg ggt gcg ggc gcg gcc tcg gtc tgc ctg gcg ctg ttg tgg ctg ctg ggg ctg	
241	A P G A G A A S V V S L A L L W L L L G L	
301	ccg tgg acc tgg agc gcg gca gcg gtc ctc ggc gtc tac gtc ggc agc ggc ggc tgg cgc	
361	P W T W S A A A A L G V Y V G S G G W R	
421	ttc ctg cgc atc gtc tgc aag acc gcg agg cga gac ctc ttc ggt ctc tct gtc ctg atc	
481	F L R I V C K T A R R D L F G L S V L I	
541	cgc gtg cgc ctg gag ctg cgg cac cag cgt gcc ggc cac acc atc ccg cgc atc ttc	
601	R V R L E L R R H Q R A G H T I P R I F	
661	cag gcg gta gtg cag cga cag ccc gag cgc ctg gcg gtg gat gcc ggc acc ggc gag	
721	Q A V V Q R Q P E R L A L V D A G T G E	
781	tgc tgg acc ttc gcg cag ctg gac gcc tac tcc aat gcc gta gcc aac ctc ttc cgc cag	
841	C W T F A Q L D A Y S N A V A N L F R Q	
901	ctg ggc ttc gcg ccg ggc gac gtg gtc gcc atc ttc ctg gag ggc ccg gag ttc gtg	
961	L G F A P G D V V A I F L E G R P E F V	
1021	ggg ctg tgg ggc ctg gcc aag gcg ggc atg gag gcc gcg ctg ctc aac gtg aac ctg	
1081	G L W L G L A K A G M E A A L L N V N L	

FIG. 26A

661 cgg cgc gag ccc ctg gcc ttc tgc ctg ggc acc tgc ggc gct aag gcc ctg atc ttt gga  
 R R E P L A F C L G T S G A K A L I F G  
 721 gga gaa atg gtg gcg gct gcc gaa gtg agc ggc cat ctg ggg aaa agt ttg atc aag  
 G E M V A A V A E V S G H L G K S L I K  
 781 ttc tgc tct gga gac ttg ggg ccc gag ggc atc ttg ccg gac acc cac ctg gac ccg  
 F C S G D L G P E G I L P D T H L L D P  
 841 ctg ctg aag gag gcc tct act gcc ccc ttg gca cag atc ccc agc aag gcc atg gac gat  
 L L K E A S T A P L A Q I P S K G M D D  
 901 cgt ctt ttc tac atc tac acg tgc ggg acc acc ggg ctg ccc aag gct gcc att gtc gtg  
 R L F Y I Y T S G T T G L P K A A I V V  
 961 cac agc agg tac tac cgc atg gca gcc ttc ggc cac ccc tac cgc atg cag gcg gct  
 H S R Y Y R M A A F G H H A Y R M Q A A  
 1021 gac gtg ctc tat gac tgc ctg ccc ctg tac cac tgc gca gga aac atc atc ggc gtg ggg  
 D V L Y D C L P L Y H S A G N I I G V G  
 1081 cag tgt ctc atc tat ggg ctg aca gtc gtc ctc cgc aag aaa ttc tgc gcc agc cgc ttc  
 Q C L I Y G L T V V L R K K F S A S R F  
 1141 tgg gac gac tgc atc aag tac aac tgc acg gtg gtt cag tac atc ggg gag atc tgc cgc  
 W D D C I K Y N C T V V Q Y I G E I C R  
 1201 tac ctg ctg aag cag ccg gtg cgc gag gcg gag agg cga cac cgc gtg cgc ctg gcg gtg  
 Y L L K Q P V R E A E R R H R V R L A V

FIG. 26B

1261 ggg aac ggg ctg cgt cct gcc atc tgg gag gag ttc acg gag cgc ttc ggc gta cgc caa  
 G N G L R P A I W E E F T E R F G V R Q  
 1321 atc ggg gag ttc tac ggc gcc acc gag tgc aac tgc agc att gcc aac atg gac ggc aag  
 I G E F Y G A T E C N C S I A N M D G K  
 1381 gtc ggc tcc tgt ggt ttc aac agc cgc atc ctg ccc cac gtg tac ccc, atc cgg ctg gtg  
 V G S C G F N S R I L P H V Y P I R L V  
 1441 aag gtc aat gag gac aca atg gag ctg ctg cgg gat gcc cag ggc ctg tgc atc ccc tgc  
 K V N E D T M E L L R D A Q G L C I P C  
 1501 cag gcc ggg gag cct ggc ctg ctt gtg ggt cag atc aac caa cag gac cgg ctg cgc cgc  
 Q A G E P G L L V G Q I N Q Q D P L R R  
 1561 ttc gat ggc tat gtc agc gag agc gcc acc agc aag aag atc gcc cac agc gtc ttc agc  
 F D G Y V S E S A T S K K I A H S V F S  
 1621 aag ggc gac agc gcc tac ctc tca ggt gac gtg cta gtg atg gat gag ctg ggc tac atg  
 K G D S A Y L S G D V L V M D E L G Y M  
 1681 tac ttc cgg gac cgt agc ggg gac acc ttc cgc tgg cga ggg gag aac gtc tcc acc acc  
 Y F R D R S G D T F R W R G E N V S T T  
 1741 gag gtg gag ggc gtg ctg agc cgc ctg ctg ggc cag aca gac gtg gcc gtc tat ggg gtg  
 E V E G V L S R L L G Q T D V A V Y G V  
 1801 gct gtt cca gga gtg gag ggt aag gca ggg atg gcg gcc gtc gca gac ccc cac agc ctg  
 A V P G V E G K A G M A A V A D P H S L  
 1861 ctg gac ccc aac gcg ata tac cag gag ctg cag aag gtg ctg gca ccc tat gcc cgg ccc  
 L D P N A I Y Q E L Q K V L A P Y A R P

FIG. 26C

```

1921  atc ttc ctg cgc ctc ctg ccc cag gtg gac acc aca ggc acc ttc aag atc cag aag acg
      I F L R L L L P Q V D T T G T F K I Q K T
1981  agg ctg cag cga gag ggc ttt gac cca cgc cag acc tca gac cgg ctc ttc ttc ctg gac
      R L Q R E G F D P R Q T S D R L F L D
2041  ctg aag cag ggc cac tac ctg ccc tta aat gag gca gtc tac act cgc atc tgc tgc ggc
      L K Q G H Y L P L N E A V Y T R I C S G
2101  gcc ttc gcc ctc tga agc tgt tcc tct act ggc cac aaa ctc tgg gcc tgg tgg gag agg
      A F A L *
2161  cca gct tga gcc aga cag cgc tgc cca ggg gtg gcc gcc tag tac aca ccc acc tgg ccg
2221  agc tgt acc tgg cac ggc cca tcc tgg act gag aaa ctg gaa cct cag agg aac ccg tgc
2281  ctc tct gct gcc ttg gtg ccc ctg tgt ctg cct cct ctc cct gct ttt cag cct ctg tct
2341  cct tcc atc cct gtc cct gtc tgg cct taa ctc ttc cct ctc ttt ctt ttt ctt tct
2401  ttc ttt ttt aag ata gag tct cac tct gct gcc cgg gct aga gtg cag tgg tgg gat
2461  ctc ggc tca ctg caa cct ctg cct cct ggg gtt caa gtg atc ctc. cca cct cag cct cct

```

FIG. 26D

2521 gag tag ctg gga tta cag gca ccc gcc acc aog tcc agc taa ttt tta tat ttt tag tag  
 2581 aga cgg ggt ttc acc atg ttg gtc agg ctg gtc ttg aac tcc tga cct cag gtg atc cgc  
 2641 tgg cct egg cct ccc aga gtg ctg gga tta tag gag tga gcc tct ggc ccg gcc ttt cct  
 2701 ttt tcc tct cct ctc ctg ccg aga gtg gaa cac acg tgt cct ggg agc tgc atc ttg tgt  
 2761 agg gtc cag ctg ctt ttg ggg act gca gga atc tcc cct ggg ccc tgg act cgg act  
 2821 ggg gcc tcc cca cct ccc cct ttg ggg act cgg ctg tgc ctt acg gag ccc caa tcc agg cct cct gtg  
 2881 gct gtt ggg ttc cag atg ctg cag ctc cat gtg act tcc aag cag gcc ctc cgc cct ccc  
 2941 tgc tga atg gag gag ccg ggg gtc ccc cag gcc aac tgg aaa atc tcc cag gct agg cca  
 3001 att gcc ttt tgc act tcc ccc ggc ttc ctg tca cat ttc ccc agc ccc acc ttc ccc tcc tga  
 3061 tgc cct gaa agc ttc cgg aat tga ctg tga cca ctt gga tgt cac cac tgt cag ccc ctg  
 3121 cct tga tgt ccc cat tta gcc atc tcc atg gag ctc ctg ctg gag ggc cct gaa ccc tgc  
 3181 act gcg tgg tgg ctg ccc agc cag ctg cct cct gtc ctg gga gcc ctc ctg ggt gtc ctc  
 3241 atc tgg tgt gtc ttc tgg agg gtc cca cag gag agg cag gag ggt cag ggg agg tct  
 3301 cct gcc ggg ggt tgg cct ctc aag cct cag ggg ttc tag cct gtt gaa tat acc cca cct  
 3361 ggt ggg tgg ccc ctc cga tgt ccc cac tga tgg ctc tga cac cgt gtt ggt ggc gat gtc  
 3421 cca gac aat ccc acc agg acg gcc cag aca tcc cta ctg gct tgg ctg gtg gct cat ctc  
 3481 gaa cat cca cgc cag cct ttc tgg ggc cca ccc agg ccg cct gtc cgt ctg tcc tcc  
 3541 ctc cag cag ccc ctg gcc cct gga gtg ggc cca tgg caa gag aca ccg tgg cgt  
 3601 ctc atg tga act ttc ctg ggc act gtg gtt tta ttt cct aat tga ttt aag aaa taa acc  
 3661 tga aga ccg tct ggt gaa aaa aaa agg gcg gcc gc

FIG. 26E

1 cga ccc acg cgt ccg ggc ggg cgg ggc gcg gcg ggc gct ggc ggg gcg gcc ggc  
 61 cca tgc agg gcg cag agc cta aac cct gct gag acc cgg ctc cgt gcg tcc agg ggc  
 121 ggc taa tgc ccc tca cgc tgc cta cgc tgc tgc aac cgg gcc gca tct gga cgg ggc gcc  
 181 gcg cgg cgg agc cga cgc cgg gcc aca atg ctg ctt gga gcc tct ctg gtc ggg gtcg ctg  
 M L L G A S L V G V L  
 241 ctg ttc tcc aag ctg gtc aaa ctg ccc tgg acc cag gtc gga ttc tcc ctg ttg ttc  
 L F S K L V L K L P W T Q V G F S L L F  
 301 ctc tac ttg gga tct ggc ggc tgg cgc ttc atc cgg gtc ttc atc aag acc atc agg cgc  
 L Y L G S G G W R F I R V F I K T I R R  
 361 gat atc ttt ggc ggc ctg gtc ctc ctg aag gtc aag gca aag gtc cga cag tgc ctg cag  
 D I F G G L V L L K V K A K V R Q C L Q  
 421 gag cgg cgg aca gtc ccc att ttg ttt gcc tct acc gtt cgg cgc cac ccc gac aag acg  
 E R R T V P I L F A S T V R R H P D K T  
 481 gcc ctg atc ttc gag ggc aca gat acc cac tgg acc ttc cgc cag ctg gat gag tac tca  
 A L I F E G T D T H W T F R Q L D E Y S  
 541 agc agt gta gcc aac ttc ctg cag gcc cgg ggc ctg gcc tgg ggc gat gtc gct gcc atc  
 S S V A N F L Q A R G L A S G D V A A I  
 601 ttc atg gag aac cgc aat gag ttc gtc ggc cta tgg ctg ggc atg gcc aag ctc ggt gtc  
 F M E N R N E F V G L W L G M A K L G V

FIG. 27A

661 gag gca gcc ctc atc aac acc aac ctg cgg gat gct ctg ctc cac tgc ctc acc acc  
 E A A L I N T N L R R D A L L H C L T T  
 721 tcg cgc gca cgg gcc ctt gtc ttt ggc agc gaa atg gcc tca gcc atc tgt gag gtc cat  
 S R A R A L V F G S E M A S A I C E V H  
 781 gcc agc ctg gac ccc tcg ctc agc ctc ttc tgc tct ggc tcc tgg gag ccc ggt gcg gtg  
 A S L D P S L S L F C S G S W E P G A V  
 841 cct cca agc aca gaa cac ctg gac cct ctg ctg aaa gat gct ccc aag cac ctt ccc agt  
 P P S T E H L D P L L K D A P K H L P S  
 901 tgc cct gac aag ggc ttc aca gat aaa ctg ttc tac atc tcc ggc acc aca ggg  
 C P D K G F T D K L F Y I Y T S G T T G  
 961 ctg ccc aag gcc gcc atc gtg gtg cac agc agc tat tac cgc atg gct gcc ctg gtg tac  
 L P K A A I V V H S R Y Y R M A A L V Y  
 1021 tat gga ttc cgc atg cgg ccc aac gac atc gtc tat gac tgc ctc ccc ctc tac cac tc  
 Y G F R M R P N D I V Y D C L P L Y H S  
 1081 gca gga aac atc gtg gga atc ggc cag tgc ctg ctg cat ggc atg acg gtg gtg att cc  
 A G N I V G I G Q C L L H G M T V V I R  
 1141 aag aag ttc tca gcc tcc cgg ttc tgg gac gat tgt atc aag tac aac tgc acg att gt  
 K K F S A S R F W D C I K Y N C T I V  
 1201 cag tac att ggt gaa ctg tgc cgc tac ctc aac cag cca ccg cgg gag gca gaa ac  
 Q Y I G E L C R Y L L N Q P P R E A E N

FIG. 27B

1261 cag cac cag gtt cgc atg gca cta ggc aat ggc ctc cgg cag tcc atc tgg acc aac ttt  
 Q H Q V R M A L G N G L R Q S I W T N F  
 1321 tcc agc cgc ttc cac ata ccc cag gtg gct gag ttc tac ggg gcc aca gag tgc aac tgt  
 S S R F H I P Q V A E F Y G A T E C N C  
 1381 agc ctg ggc aac ttc gac agc cag gtg ggg gcc tgt ttc aat agc cgc atc ctg tcc  
 S L G N F D S Q V G A C G F N S R I L S  
 1441 ttc gtg tac ccc atc cgg ttg gta cgt gtc aac gag gac acc atg gag ctg atc cgg ggg  
 F V Y P I R L V R V N E D T M E L I R G  
 1501 ccc gac ggc gtc tgc att ccc tgc cag cca ggt gag ccg ggc cag ctg gtg ggc cgc atc  
 P D G V C I P C Q P G E P G Q L V G R I  
 1561 atc cag aaa gac ccc ctg cgc cgc ttc gat ggc tac ctg aac cag gcc gcc aac aac aag  
 I Q K D P L R R F D G Y L N Q G A N N K  
 1621 aag att gcc aag gat gtc ttc aag aag ggc gac cag gcc tac ctt act ggt gat gtg ct~  
 K I A K D V F K K G D Q A Y L T G D V L  
 1681 gtg atg gac gag ctg ggc tac ctg tac ttc cga gac cgc act ggg gac acg ttc cgc tg  
 V M D E L G Y L Y F R D R T G D T F R W  
 1741 aaa ggt gag aac gtg tcc acc acc gag gtg gaa ggc aca ctc agc cgc ctg gac at  
 K G E N V S T T E V E G T L S R L L D M  
 1801 gct gac gtg gcc gtg tat ggt gtc gag gtg cca gga acc gag gcc ggc gga atg gc  
 A D V A V Y G V E V P G T E G R A G M A  
 1861 gct gtg gcc agc ccc act ggc aac tgt gac ctg gag cgc ttt gct cag gtc ttg gag aag  
 A V A S P T G N C D L E R F A Q V L E K

FIG. 27C



1921 gaa ctg ccc ctg tat gcg cgc ccc atc ttc ctg cgc ctc ctg cct gag ctg cac aaa aca  
 E L P L Y A R P I F L R L L P E L H K T  
 1981 gga acc tac aag ttc cag aag aca gag cta cgg aag gag ggc ttt gac ccg gct att gtg  
 G T Y K F Q K T E L R K E G F D P A I V  
 2041 aaa gac ccg ctg ttc tat cta gat gcc cag aag ggc cgc tac gtc ccg ctg gac caa gag  
 K D P L F Y L D A Q K G R Y V P L D Q E  
 2101 gcc tac agc cgc atc cag gca ggc gag aag ctg tga ttc ccc cca tcc ctc tga ggg  
 A Y S R I Q A G E K L \*  
 2161 ccg gcg gat gct gga tcc gga gcc cca ggt tcc gcc cca gag cgg tcc tgg aca agg cca  
 2221 gac caa agc aag cag ggc ctg gca cct cca tcc tga ggt gct gcc cct cca tcc aaa act  
 2281 gcc aag tga ctc att gcc ttc cca acc ctt cca gag gct ttc tgt gaa agt ctc atg tcc  
 2341 aag ttc cgt ctt ctg ggc tgg gca gcc cct ctg gtt ccc agg ctg aga ctg acg ggt ttt  
 2401 ctc agg atg atg tct tgg gtg agg gta ggg aga gga cca ggc gtc acc gag ccc ttc cca  
 2461 gag agc agg gag ctt ata aat gga acc aga gca gta gtc ccc aga ctc agg aag tca a  
 2521 gag tgg gca ggg aca gtg gta gca tcc atc tgg tgg cca aag aga atc gta gcc cca c  
 2581 ctg ccc aag ttc act ggg ctc cac ccc cac ctc cag gag ggg agg aga gga cct gac a  
 2641 tgt agg tgg ccc ctg atg ccc cat cta cag cag gtc agg acc acg ccc ctg gcc t  
 2701 ccc cac tcc ccc atc ctc ctc cct ggg tgg ctg cct gat tat ccc tca ggc agg gcc t  
 2761 cag tcc ttg tgg gtc tgt gtc acc tcc atc tca gtc ttg gcc tgg cta tga ggg gag c  
 2821 gaa tgg gag agg ggg ctc agg ggc caa taa act ctg cct tga gtc ctc cta aaa aaa c  
 2881 aaa aaa aaa aaa aaa aaa ggg cgg ccg c

FIG. 27D

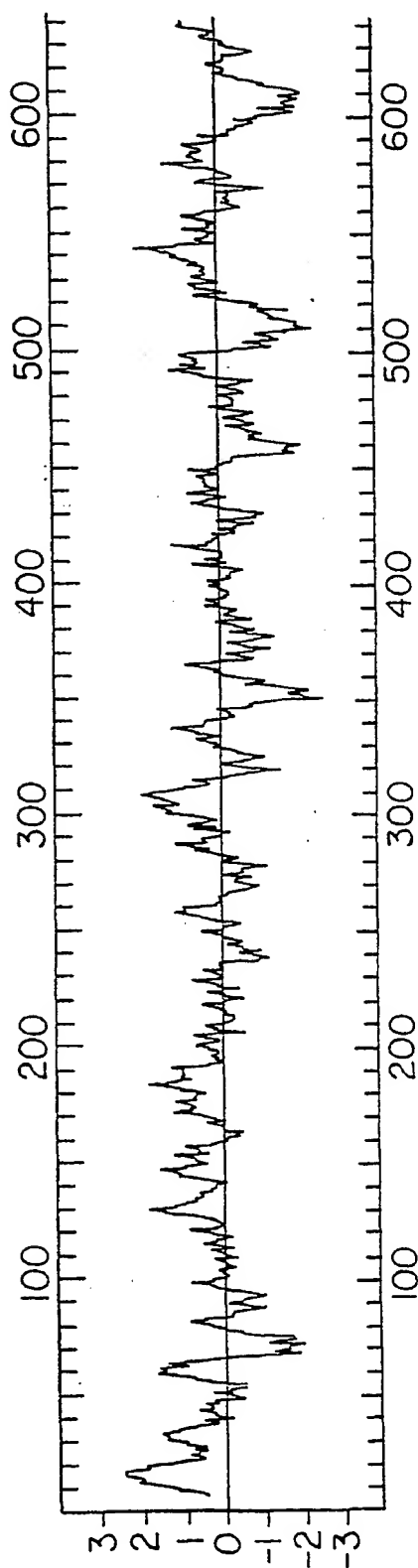


FIG. 28A

		<u>n</u>	<u>n%</u>	<u>MW</u>	<u>MW%</u>
A ala	Alanine	64	9.9	4546	6.4
C cys	cysteine	15	2.3	1545	2.2
D asp	aspartic acid	30	4.6	3450	4.9
E glu	glutamic acid	31	4.8	4000	5.6
F phe	phenylalanine	29	4.5	4264	6.0
G gly	glycine	63	9.8	3592	5.1
H his	histidine	13	2.0	1781	2.5
I ile	isoleucine	29	4.5	3279	4.6
K lys	lysine	22	3.4	2818	4.0
L leu	leucine	77	11.9	8707	12.3
M met	methionine	11	1.7	1441	2.0
N asn	asparagine	15	2.3	1710	2.4
P pro	proline	29	4.5	2814	4.0
Q gln	glutamine	25	3.9	3201	4.5
R arg	arginine	49	7.6	7648	10.8
S ser	serine	33	5.1	2872	4.0
T thr	threonine	27	4.2	2728	3.8
V val	valine	51	7.9	5052	7.1
W trp	tryptophan	9	1.4	1674	2.4
X ukw	unknown	--	--		
Y tyr	tyrosine	24	3.7	3913	5.5
Z ---	STOP				

FIG. 28B

SUBSTITUTE SHEET (RULE 26)

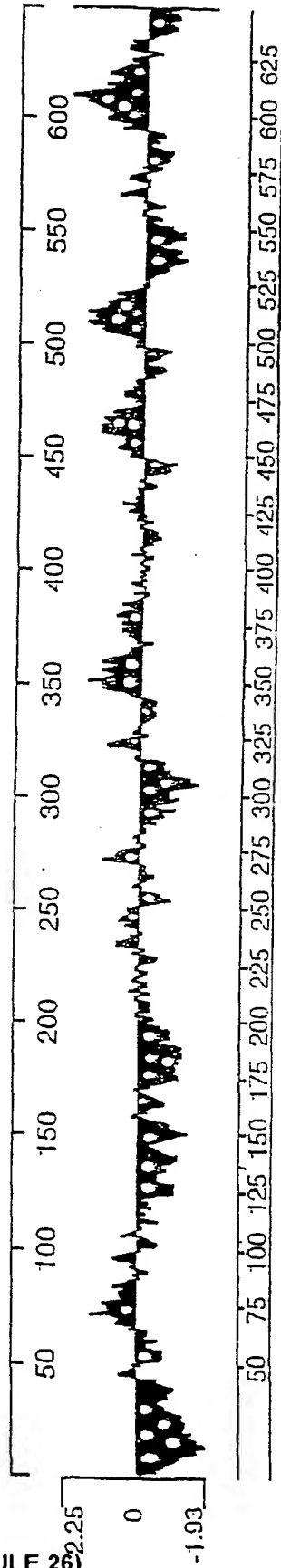


FIG. 28C

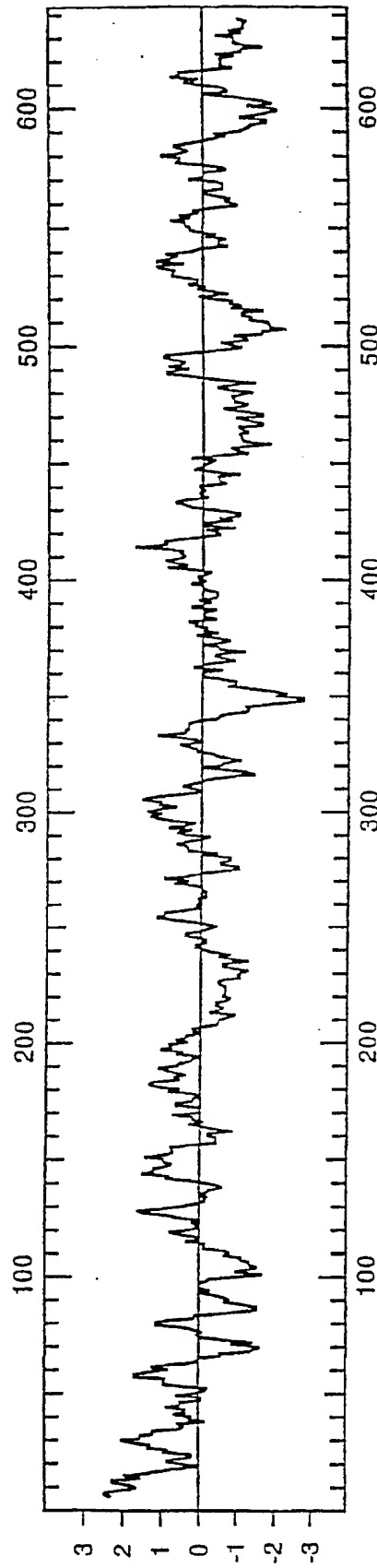


FIG. 29A

		<u>n</u>	<u>n%</u>	<u>MW</u>	<u>MW%</u>
A ala	alanine	46	7.2	3267	4.5
C cys	cysteine	16	2.5	1648	2.3
D asp	aspartic acid	33	5.1	3795	5.3
E glu	glutamic acid	33	5.1	4258	5.9
F phe	phenylalanine	34	5.3	5000	6.9
G gly	glycine	54	8.4	3079	4.3
H his	histidine	12	1.9	1644	2.3
I ile	isoleucine	30	4.7	3392	4.7
K lys	lysine	31	4.8	3970	5.5
L leu	leucine	76	11.8	8594	11.9
M met	methionine	12	1.9	1572	2.2
N asn	asparagine	21	3.3	2394	3.3
P pro	proline	31	4.8	3008	4.2
Q gln	glutamine	23	3.6	2945	4.1
R arg	arginine	45	7.0	7024	9.8
S ser	serine	35	5.4	3046	4.2
T thr	threonine	32	5.0	3233	4.5
V val	valine	46	7.2	4557	6.3
W trp	tryptophan	8	1.2	1488	2.1
X ukw	unknown	--	--		
Y tyr	tyrosine	25	3.9	4076	5.7
Z ---	STOP				

FIG. 29B

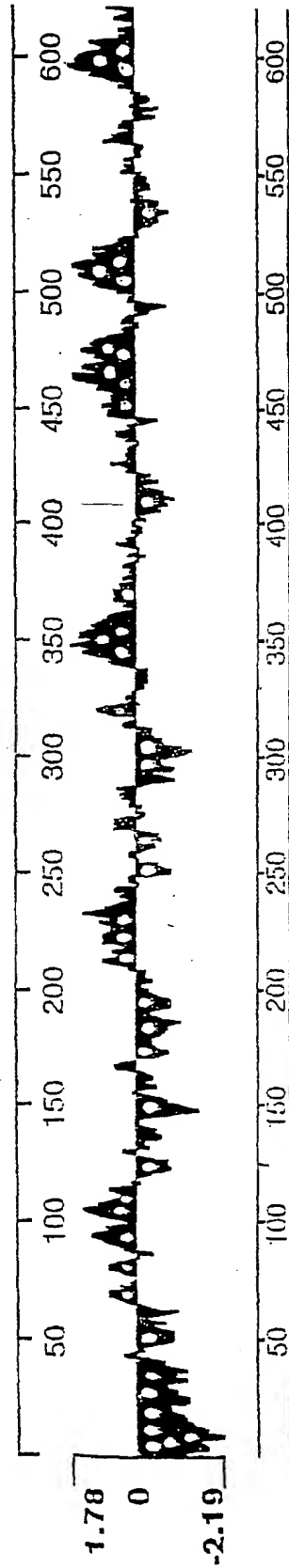


FIG. 29C

1	A	T	G	C	G	G	C	T	C	C	G	G	T	G	C	G	G	hFATP1
1	A	T	G	C	G	G	C	T	C	C	T	G	G	A	G	C	A	mFATP1
21	C	G	C	G	C	C	T	C	G	G	T	G	G	T	C	T	C	hFATP1
21	A	A	C	A	G	C	C	T	C	T	G	T	G	G	C	C	A	mFATP1
41	T	G	G	C	G	C	T	G	T	G	T	G	G	C	T	G	C	hFATP1
41	T	G	G	C	G	C	T	G	T	G	T	G	G	T	T	C	T	mFATP1
61	G	G	G	C	T	G	C	C	G	T	G	G	A	C	C	T	G	hFATP1
61	G	G	A	C	T	T	C	C	G	T	G	G	A	C	C	T	G	mFATP1
81	C	G	C	G	G	C	A	G	C	G	G	C	C	C	T	C	G	hFATP1
81	C	G	C	G	G	C	G	G	C	G	C	C	G	T	T	C	T	mFATP1
101	T	G	T	A	C	G	T	G	G	C	A	G	C	G	G	C	G	hFATP1
101	T	G	T	A	C	G	T	G	G	T	G	G	C	G	C	G	C	mFATP1
121	T	G	G	C	G	C	T	T	C	C	T	G	C	G	C	A	T	hFATP1
121	T	G	G	C	G	C	T	T	T	C	C	T	G	C	G	T	A	mFATP1
141	C	T	G	C	A	A	G	A	C	C	G	C	G	A	G	G	C	hFATP1
141	C	T	G	C	A	A	G	A	C	G	G	C	G	A	G	G	C	mFATP1
161	A	C	C	T	C	T	T	C	G	G	T	C	C	T	C	T	G	hFATP1
161	A	C	C	T	C	T	T	T	G	G	C	C	T	C	T	C	T	mFATP1

FIG. 30A



181	C	T	G	A	T	C	C	G	C	G	T	G	C	C	T	G	G	A	hFATP1
181	C	T	G	A	T	T	C	C	T	G	T	T	C	G	G	C	T	A	mFATP1
201	G	C	T	G	C	G	G	C	G	G	C	A	C	C	A	G	C	G	hFATP1
201	G	C	T	G	C	G	A	C	C	A	C	A	C	C	G	G	C	A	mFATP1
221	C	C	G	G	C	C	A	C	A	C	C	A	T	C	C	C	G	C	hFATP1
221	C	A	G	G	A	G	A	C	A	C	G	A	T	C	C	C	G	T	mFATP1
241	A	T	C	T	T	T	C	A	G	G	C	G	G	T	A	G	T	G	hFATP1
241	A	T	C	T	T	C	C	A	G	G	C	T	G	T	G	C	C	G	mFATP1
261	G	C	G	A	C	A	G	C	C	C	G	A	G	C	G	C	C	T	hFATP1
261	G	C	G	A	C	A	A	C	C	A	G	A	G	C	G	C	C	T	mFATP1
281	C	G	C	T	G	G	T	G	G	A	T	G	C	C	G	G	A	C	hFATP1
281	C	A	C	T	G	G	T	G	G	A	C	G	C	C	A	G	T	A	mFATP1
301	G	G	C	G	A	G	T	G	C	T	G	G	A	C	C	T	T	G	hFATP1
301	G	G	T	A	T	A	T	G	C	T	G	G	A	C	C	T	T	C	mFATP1
321	G	C	A	G	C	T	G	G	A	C	G	C	T	A	C	T	C	C	hFATP1
321	A	C	A	G	C	T	G	G	A	C	A	C	C	T	A	C	T	C	mFATP1
341	A	T	G	C	G	G	T	A	G	C	C	A	A	C	C	T	C	T	hFATP1
341	A	T	G	C	T	G	T	A	G	C	C	A	A	C	C	T	G	T	mFATP1

FIG. 30B

721	G A	C	G A T C G	T	C T	T	T T	C	T A C A T	hFATP1
721	G A	T	G A T C G	G	C T	G	T T	T	T A C A T	mFATP1
741	C T A	C A C	G T C	G	G G G A C C A C C G					hFATP1
741	C T A	T A C	T T C	T	G G G G A C C A C C G					mFATP1
761	G G C T	G C C	C A A G G C T G C C A T T							hFATP1
761	G G C T	T C C	T A A G G C T G C C A T T							mFATP1
781	G T	C G T G C A C A G C A G G T A C T A								hFATP1
781	G T	G T G C A C A G C A G G T A C T A								mFATP1
801	C C G C A T	G G C A	G C C C T T C	G G C C						hFATP1
801	C C G C A T	T G C T	G C C C T T	T G G C C						mFATP1
821	A C C A	C G C C T A C	C G C A T G C C A G							hFATP1
821	A C C A	T C C T A C	A G C A T G C C T							mFATP1
841	G C	G G C T G A C	G T G C C T C T A T G A							hFATP1
841	G C	C G C C G A T	G T G C C T C T A T G A							mFATP1
861	C T G C C C T G C C	C C T G	T A C C A C T							hFATP1
861	C T G C C C T G C C	A C T C	T A C C A C T							mFATP1
881	C G G C A G G A	A A C A T C A T	C G G C							hFATP1
881	C T G C A G G A	A A C A T C A T	G G G T							mFATP1

FIG. 30C

901	G T G G G G C A G T G	T C T C A T C T A	hFATP1
901	G T G G G G C A G T G	C G T C A T C T A	mFATP1
921	T G G G C T G A C A G T C	C G T C C T C C	hFATP1
921	C G G G T T G A C A G T C	G G T A C T G C	mFATP1
941	G C A A G A A A T T C T C	G G C C A G C	hFATP1
941	G C A A G A A G T T C T C	C G C C A G C	mFATP1
961	C G C T T C T G G G A C	G A C T G C A T	hFATP1
961	C G C T T C T G G G A T	G A C T G T G T	mFATP1
981	C A A G T A C A A C T G C A	C G G T G G T G G	hFATP1
981	C A A G T A C A A T T G C A	C G G T G G T A G	mFATP1
1001	T T C A G T A C A T C G G G	G A G A T C	hFATP1
1001	T G C A G T A C A T A G G T	G A A T C	mFATP1
1021	T G C C G C C T A C C C T G C	T G A A G C A	hFATP1
1021	T G C C G C C T A C C C T G C	T G A A G C A	mFATP1
1041	G C C G G T G C G C G A G G	C G G A G A	hFATP1
1041	G C C G G T T C G C G A C G	T G G A G C	mFATP1
1061	G C G A C A C C G C G T G C	G C C C T G	hFATP1
1061	A G C G A C A C C G C G T G C	G C C C T G	mFATP1

FIG. 30D

1081	G C	G G T G G G	G A A	C G G G C T G C G	hFATP1
1081	G C	C G T G G G	T A A	T G G G C T G C G	mFATP1
1101	T C C	T G C C A T C T	G G G A G G A G T	hFATP1	
1101	G C C	A G C C A T C T	G G G A G G A G T	mFATP1	
1121	T C A C G G	G A G C G C T T	C G G C G T A	hFATP1	
1121	T C A C G G	C A G C G C T T	C G G C G T G	mFATP1	
1141	C G C C A A	A T C G G G G	G A G T T C T A	hFATP1	
1141	C C A C A G	A T C G G C C	G A G T T C T A	mFATP1	
1161	C G G C G C C	C A C C G A G T	G C A A C T	hFATP1	
1161	C G G C G C C	T A C C G A G T	G C A A C T	mFATP1	
1181	G C A G C A T	T G C C C A A C	A T G G A C	hFATP1	
1181	G C A G C A T	T G C C C A A C	A T G G A C	mFATP1	
1201	G G C A A G G T	C G G C T C C T	G T G G	hFATP1	
1201	G G C A A G G T	C G G C T C C T	G T G G	mFATP1	
1221	T T T C A A C A	G C C C G C A T	C C T G C	hFATP1	
1221	C T T C A A C A	G C C C G C A T	C C T G C	mFATP1	
1241	C C C A C G T G T A	C C C C A T C C G G	hFATP1		
1241	C G C A C G T G T A	C C C C A T C C G T	mFATP1		

FIG. 30E

1261	C	T	G	G	T	G	A	A	G	G	T	C	A	A	T	G	A	G	G	A	hFATP1	
1261	C	T	G	G	T	C	A	A	G	G	T	C	A	A	T	G	A	G	G	A	mFATP1	
1281	C	A	C	A	A	T	G	G	A	G	C	T	G	C	G	G	G	G	G	hFATP1		
1281	C	A	C	G	A	T	G	G	A	G	C	C	A	C	T	G	C	G	G	mFATP1		
1301	A	T	G	C	C	C	A	G	G	G	C	C	T	C	T	G	C	A	T	C	hFATP1	
1301	A	C	T	C	C	G	A	G	G	G	C	C	T	C	T	G	C	A	T	C	mFATP1	
1321	C	C	C	T	G	C	C	A	G	C	C	G	G	G	G	A	G	C	C	hFATP1		
1321	C	C	G	T	G	C	C	A	G	C	C	G	G	G	G	A	A	C	C	mFATP1		
1341	T	G	G	C	C	T	C	C	T	C	T	G	T	G	G	G	T	C	A	G	A	hFATP1
1341	C	G	G	C	C	T	T	C	T	C	T	G	T	G	G	C	C	A	G	A	mFATP1	
1361	T	C	A	A	C	C	A	A	C	A	G	C	A	G	G	A	C	C	C	T	G	hFATP1
1361	T	C	A	A	C	C	A	G	C	A	G	C	A	G	G	A	C	C	C	T	G	mFATP1
1381	C	G	C	C	G	C	T	T	C	G	A	T	G	G	C	T	A	T	G	T	hFATP1	
1381	C	G	G	C	C	T	T	T	C	G	A	T	G	G	T	T	A	T	G	T	mFATP1	
1401	C	A	G	C	G	A	G	A	G	C	G	C	C	A	C	C	A	G	C	A	hFATP1	
1401	T	A	G	T	G	A	C	A	G	T	G	C	C	A	C	C	A	A	C	A	mFATP1	
1421	A	G	A	A	G	A	T	C	G	C	C	C	A	C	A	G	C	C	T	C	hFATP1	
1421	A	G	A	A	G	A	T	T	G	C	C	C	A	C	A	G	C	C	T	T	mFATP1	

FIG. 30F

1441	T	T	C	A	G	C	A	A	G	G	G	C	G	A	C	A	G	C	G	C	hFATP1
1441	T	T	C	C	G	A	A	A	G	G	G	C	G	A	T	A	G	C	G	C	mFATP1
1461	C	T	A	C	C	T	C	T	C	A	G	G	T	G	A	C	G	T	G	C	hFATP1
1461	C	T	A	C	C	T	C	T	C	A	G	G	T	G	A	C	G	T	G	C	mFATP1
1481	T	A	G	T	G	A	T	G	G	A	T	G	A	G	C	T	G	G	G	C	hFATP1
1481	T	A	G	T	G	A	T	G	G	A	C	G	A	G	C	T	G	G	G	C	mFATP1
1501	T	A	C	A	T	G	T	A	C	T	T	C	C	G	G	A	C	C	G	hFATP1	
1501	T	A	C	A	T	G	T	A	T	T	T	C	C	G	T	G	A	C	C	G	mFATP1
1521	T	A	G	C	G	G	G	A	C	A	C	C	T	T	C	C	G	C	T	hFATP1	
1521	C	A	G	C	G	G	G	A	C	A	C	C	T	T	C	C	G	C	T	mFATP1	
1541	G	G	C	G	A	G	G	G	A	G	A	A	C	G	T	C	T	C	C	hFATP1	
1541	G	G	C	G	C	G	G	G	A	G	A	A	C	G	T	G	T	C	C	mFATP1	
1561	A	C	C	A	C	C	G	A	G	G	T	G	G	A	G	G	C	G	T	hFATP1	
1561	A	C	C	A	C	G	G	A	G	G	T	G	G	A	A	G	C	C	G	mFATP1	
1581	G	C	T	G	A	G	C	C	G	C	C	T	G	C	T	G	G	G	C	C	hFATP1
1581	G	C	T	G	A	G	C	C	G	C	C	T	A	C	T	G	G	G	C	C	mFATP1
1601	A	G	A	C	A	G	A	C	C	T	G	G	C	C	G	T	C	T	A	T	hFATP1
1601	A	G	A	C	G	G	A	C	C	T	G	G	C	C	T	G	T	C	T	A	mFATP1

FIG. 30G

1621	G G G G T G G C C T G T	T C C A G G A G T	hFATP1
1621	G G G G T G G C C T G T	C C C A G G A G T	mFATP1
1641	G G A G G G T A A G G C A	G G G A T G G	hFATP1
1641	G G A G G G G A A A G C T	G G C A T G G	mFATP1
1661	C G G C C G T C G C A G A C	C C C C C A C	hFATP1
1661	C A G C C A T C G C A G A C	T C C C C A C	mFATP1
1681	A G C C C T G C T G G A C C C	C A A C G C	hFATP1
1681	A G C C C A G T T G G A C C C	T A A C T C	mFATP1
1701	G A T A T A C C A G G A G C	T G C A G A	hFATP1
1701	A A T G T A C C A G G A A T	T A C A G A	mFATP1
1721	A G G T G C T G G C A C C C	T A T G C C	hFATP1
1721	A G G T T C T T G C A T C C	T A T G C T	mFATP1
1741	C G G C C C A T C T T C C T	G C C G C C T	hFATP1
1741	C G G C C C A T C T T C C T	G C C G C C T	mFATP1
1761	C C T G C C C C A G G T G G A	C A C C A	hFATP1
1761	T C T G C C C C A G G T G G A	T A C C A	mFATP1
1781	C A G G C A C C T T C A A G A	T C C A G	hFATP1
1781	C A G G C A C C T T C A A G A	T C C A G	mFATP1

FIG. 30H

1801	A	A	G	A	C	G	A	G	G	C	T	G	C	A	G	C	G	A	G	A	hFATP1
1801	A	A	G	A	C	C	C	G	G	C	T	G	C	A	G	C	G	T	G	A	mFATP1
1821	G	G	G	C	T	T	T	G	A	C	C	C	A	C	G	C	C	A	G	A	hFATP1
1821	A	G	G	C	T	T	T	G	A	C	C	C	C	C	G	T	C	A	G	A	mFATP1
1841	C	C	T	C	A	G	A	C	C	G	G	C	T	C	T	T	C	T	T	C	hFATP1
1841	C	C	T	C	A	G	A	C	A	G	G	C	T	C	T	T	C	T	T	T	mFATP1
1861	C	T	G	A	C	C	T	G	A	A	G	C	A	G	G	G	C	C	A	A	hFATP1
1861	C	T	A	G	A	C	C	T	G	A	A	G	C	A	G	G	A	C	C	G	mFATP1
1881	C	T	A	C	C	T	G	C	C	C	T	A	A	A	T	G	A	G	G	G	hFATP1
1881	C	T	A	T	G	T	A	C	C	C	C	T	G	G	A	T	G	A	G	A	mFATP1
1901	C	A	G	T	C	T	A	C	A	C	T	C	G	C	A	T	C	T	G	C	hFATP1
1901	G	A	G	T	C	C	A	T	G	C	C	C	C	A	T	T	T	G	T	T	mFATP1
1921	T	C	G	G	C	G	C	C	T	T	C	G	C	C	C	T	C	T	C	G	hFATP1
1921	G	C	A	G	G	C	G	A	C	T	T	C	T	C	A	C	T	C	C	C	mFATP1
1941	A																			hFATP1	
1938																				mFATP1	

FIG. 301



1	C	T	G	T	C	T	C	C	A	A	G	C	T	G	G	C	T	G	A	A	A	C	T	G	C	C	C	hsFATP4		
1	C	T	T	G	G	G	T	C	C	A	A	G	C	T	A	G	T	G	C	T	G	A	A	G	C	T	G	C	mmFATP4	
31	T	G	G	A	C	C	C	A	G	G	T	G	G	G	A	T	T	C	T	C	C	C	T	G	T	G	T	C	hsFATP4	
31	T	G	G	A	C	C	C	A	G	G	T	G	G	G	A	T	T	C	T	C	C	C	T	G	T	G	T	C	mmFATP4	
61	C	T	C	T	A	C	T	T	G	G	G	A	T	C	T	G	G	C	T	G	G	C	C	T	G	C	T	T	C	hsFATP4
61	C	T	G	T	A	C	T	T	G	G	G	T	C	T	G	G	T	G	G	C	T	G	G	C	C	T	T	T	C	mmFATP4
91	A	T	C	C	G	G	G	T	C	T	T	C	A	T	C	A	A	G	A	C	C	A	T	C	A	G	G	C	G	hsFATP4
91	A	T	C	C	G	G	G	T	C	T	T	C	A	T	C	A	A	G	A	C	G	G	T	C	A	G	G	A	G	mmFATP4
121	G	A	T	A	T	C	T	T	G	G	C	G	G	C	C	T	G	G	T	C	C	C	T	G	A	A	G	hsFATP4		
121	G	A	T	A	T	C	T	T	G	G	T	G	G	C	A	T	G	G	T	C	C	C	T	G	A	A	G	mmFATP4		
151	G	T	G	A	A	G	G	C	A	A	A	G	G	T	G	C	G	A	C	A	G	T	G	C	C	A	G	hsFATP4		
151	G	T	G	A	A	G	A	C	C	A	A	G	G	T	G	C	G	A	C	G	T	A	C	C	T	T	C	A	mmFATP4	
181	G	A	G	C	G	G	C	G	A	C	A	G	T	G	C	C	C	C	A	T	T	T	G	T	T	T	G	C	C	hsFATP4
181	G	A	G	C	G	G	A	A	G	A	C	A	G	T	G	C	C	C	C	T	G	C	T	T	T	T	G	C	T	mmFATP4

FIG. 31A

211	T	C	T	A	C	C	G	T	T	C	G	G	C	C	A	C	C	C	C	G	A	C	A	A	G	A	C	G	hsFATP4		
211	T	C	A	A	T	G	G	T	A	C	A	G	C	C	C	A	C	C	C	G	G	A	C	A	A	G	A	C	A	mmFATP4	
241	G	C	C	C	T	G	A	T	C	T	T	C	G	A	G	G	G	C	A	C	A	G	A	T	A	C	C	A	C	hsFATP4	
241	G	C	C	C	T	G	A	T	T	T	T	C	G	A	G	G	G	C	A	C	A	G	A	C	A	C	T	C	A	C	mmFATP4
271	T	G	G	A	C	C	T	T	C	C	G	C	C	A	G	C	T	G	G	A	T	G	A	G	T	A	C	T	C	A	hsFATP4
271	T	G	G	A	C	C	T	T	C	C	G	C	C	A	G	C	T	G	G	A	T	G	A	G	T	A	C	T	C	A	mmFATP4
301	A	G	C	A	G	T	G	T	A	G	C	C	A	A	C	T	T	C	C	T	G	C	A	G	G	C	C	C	G	G	hsFATP4
301	A	G	T	A	G	T	G	T	G	G	C	C	A	A	C	T	T	C	C	T	G	C	A	G	G	C	C	C	G	G	mmFATP4
331	G	G	C	C	T	G	G	C	C	T	C	G	G	C	G	A	T	G	T	G	G	C	T	G	C	C	A	T	C	C	hsFATP4
331	G	G	C	C	T	G	G	C	C	T	C	A	G	G	C	A	A	T	G	T	A	G	T	T	G	C	C	C	T	C	mmFATP4
361	T	T	C	A	T	G	G	A	G	A	A	C	C	G	C	A	A	T	G	A	G	T	T	C	G	T	G	G	G	C	hsFATP4
361	T	T	T	A	T	G	G	A	A	A	A	C	C	G	C	A	A	T	G	A	G	T	T	T	G	T	G	G	G	T	mmFATP4
391	C	T	A	T	G	G	C	T	G	G	C	A	T	G	G	C	C	A	A	G	C	T	C	G	G	T	G	T	G	T	hsFATP4
391	C	T	G	T	G	G	C	T	A	G	G	C	A	T	G	G	C	C	A	A	G	C	T	G	G	C	C	T	G	T	mmFATP4

FIG. 31B

421	G	A	G	G	C	A	G	C	C	T	C	A	T	C	A	A	C	A	C	C	T	G	C	G	G	hsFATP4					
421	G	A	G	G	C	G	G	C	T	C	T	C	A	T	C	A	A	C	A	C	C	T	T	A	G	mmFATP4					
451	C	G	G	G	A	T	G	C	T	C	T	G	C	T	C	C	A	C	T	G	C	C	A	C	C	hsFATP4					
451	C	G	G	G	A	T	G	C	C	C	T	G	C	T	C	C	A	C	T	G	T	C	T	G	A	mmFATP4					
481	T	C	G	C	G	C	A	C	G	G	C	C	T	T	G	T	C	T	T	T	G	G	C	A	G	C	hsFATP4				
481	T	C	A	A	A	G	G	C	A	C	G	A	G	C	T	C	T	C	T	T	T	G	G	C	A	G	mmFATP4				
511	G	A	A	T	G	G	C	C	T	C	A	G	C	C	A	T	C	T	G	T	G	A	G	T	C	C	A	hsFATP4			
511	G	A	G	A	T	G	G	C	C	T	C	A	G	C	T	A	T	C	T	G	T	G	A	G	A	T	C	C	A	mmFATP4	
541	G	C	C	A	G	C	C	T	G	G	A	C	C	C	T	C	G	C	T	C	A	G	C	C	T	T	C	hsFATP4			
541	G	C	T	A	G	C	C	T	G	G	A	G	C	C	C	A	C	A	C	T	C	A	G	C	C	T	T	C	mmFATP4		
571	T	G	C	T	C	T	G	G	C	T	C	C	T	G	G	G	A	G	C	C	C	G	T	G	C	G	T	G	hsFATP4		
571	T	G	C	T	C	T	G	G	A	T	C	C	T	G	G	G	A	G	C	C	C	A	G	C	A	C	A	G	T	mmFATP4	
601	C	C	T	C	C	A	A	G	C	A	C	A	G	A	A	C	A	C	C	T	G	G	A	C	C	C	T	C	T	G	hsFATP4
601	C	C	G	T	C	A	A	G	C	A	C	A	G	A	G	C	A	T	C	T	G	G	A	C	C	C	T	C	T	T	mmFATP4

FIG. 31C

631	C	T	G	A	A	A	G	A	T	G	C	T	C	C	C	A	G	T	hsFATP4
631	C	T	G	G	A	A	G	A	T	G	C	C	G	A	A	G	C	A	mmFATP4
661	T	G	C	C	C	T	G	A	C	A	A	G	G	C	T	T	C	A	hsFATP4
661	C	A	C	C	C	A	G	A	C	A	A	G	G	T	T	T	A	C	mmFATP4
691	T	T	C	T	A	C	A	T	C	T	A	C	A	T	C	G	G	C	hsFATP4
691	T	T	C	T	A	C	A	T	C	T	A	C	A	T	C	G	G	C	mmFATP4
721	C	T	G	C	C	C	A	A	G	C	C	G	C	C	A	T	C	G	hsFATP4
721	C	T	A	C	C	C	A	A	A	G	C	T	G	C	C	A	T	C	mmFATP4
751	A	G	G	T	A	T	T	A	C	C	G	T	A	T	G	G	C	T	hsFATP4
751	A	G	G	T	A	T	T	A	T	C	C	G	T	A	T	G	G	C	mmFATP4
781	T	A	T	G	G	A	T	T	C	C	G	C	A	T	G	C	G	C	hsFATP4
781	T	A	T	G	G	A	T	T	C	C	G	C	A	T	G	C	G	C	mmFATP4
811	G	T	C	T	A	T	G	A	C	T	G	C	C	C	T	C	T	A	hsFATP4
811	G	T	C	T	A	T	G	A	C	T	G	C	C	C	T	C	T	A	mmFATP4

FIG. 31D

841	-	G	C	A	G	G	A	A	C	A	T	C	G	T	G	G	G	A	A	T	C	G	G	C	C	A	G	T	G	hsFATP4	
841	A	G	C	A	G	G	A	A	A	C	A	T	C	G	T	G	G	G	A	T	C	T	G	G	C	-	A	G	T	G	mmFATP4
870	C	C	T	G	C	T	G	C	A	T	G	G	C	A	T	G	A	C	G	T	G	G	T	G	A	T	C	C	G	hsFATP4	
870	C	T	T	A	C	T	C	A	C	G	G	C	A	T	G	A	C	T	G	T	G	G	T	G	A	T	C	C	G	mmFATP4	
900	G	A	A	G	A	G	T	T	C	T	C	A	G	C	C	T	C	C	C	G	G	T	T	C	T	G	G	G	A	hsFATP4	
900	G	A	A	G	A	G	T	T	C	T	C	A	G	C	C	T	C	C	C	G	G	T	T	C	T	G	G	G	A	mmFATP4	
930	C	G	A	T	G	T	A	T	C	A	A	G	T	A	C	A	A	C	T	G	C	A	C	G	A	T	T	G	T	hsFATP4	
930	T	G	A	T	G	T	A	T	C	A	A	G	T	A	C	A	A	C	T	G	C	A	C	A	G	T	T	G	T	mmFATP4	
960	G	C	A	G	T	A	C	A	T	T	G	G	T	G	A	A	C	T	G	T	G	C	C	G	C	T	A	C	C	T	hsFATP4
960	A	C	A	G	T	A	C	A	T	T	G	G	T	C	G	A	A	C	T	G	C	C	G	C	T	A	C	C	T	mmFATP4	
990	C	C	T	G	A	A	C	C	A	G	C	C	A	C	C	G	C	G	G	A	G	G	C	A	G	A	A	A	A	hsFATP4	
990	C	C	T	G	A	A	C	C	A	G	C	C	A	C	C	C	C	G	T	G	A	G	G	C	T	G	A	G	T	C	mmFATP4
1020	C	C	A	G	C	A	C	C	A	G	G	T	T	C	G	C	A	T	G	G	C	A	C	T	A	G	G	C	A	A	hsFATP4
1020	T	C	G	G	C	A	C	A	A	G	G	T	T	G	C	G	C	A	T	G	G	C	A	C	T	G	G	C	A	A	mmFATP4

FIG. 31E

1050	TGG	CCTCCGGGCAAGTCCCATCTGGACC	ACCTT	hsFATP4
1050	CGG	CTCCCGGCAAGTCCCATCTGGACC	GACTT	mmFATP4
1080	TCCAGCCCG	CTTCCACATACCCCAAGGTGGC		hsFATP4
1080	CTCCAGCCG	TTCCACATCCCCCAAGGTGGC		mmFATP4
1110	TGAGTTCTA	CGGGGCCACAGAGTGCACAAC TG		hsFATP4
1110	TGAGTTCTA	TGGGGGCCACAGAGTGCACAAC TG		mmFATP4
1140	TAGCCCTGGGCAACCTT	CGACAGCCAGGTGGG		hsFATP4
1140	TAGCCCTGGGCAACCTT	TGACAGCCAGGTGGG		mmFATP4
1170	GGCCCTGGTGGT	TTCAATAGCCGCATCCTGTC		hsFATP4
1170	GGCCCTGGTGGT	TTCAATAGCCGCATCCTGTC		mmFATP4
1200	CTT	CGTGTAACCCTATCCCGTTTGGGTACGTGT		hsFATP4
1200	CTT	TGTGTACCCCTATCCCGTTTGGGTACGTGT		mmFATP4
1230	CAC	GAGGACACCATGGAGCTGATCCGGGG		hsFATP4
1230	CAT	GAGGATACCATGGAGCTGATCCGGGG		mmFATP4

FIG. 31F

1260	G	C	C	C	G	A	C	G	G	C	G	T	C	T	G	C	A	T	T	C	C	C	T	G	C	C	A	C	C	hsFATP4	
1260	A	C	C	C	G	A	T	G	G	A	G	T	C	T	G	C	A	T	T	C	C	C	T	G	T	C	A	A	C	C	mmFATP4
1290	A	G	G	T	G	A	G	C	C	G	G	G	C	C	A	G	C	T	G	G	T	G	G	G	C	C	G	C	A	T	hsFATP4
1290	A	G	G	T	C	A	G	C	C	A	G	G	C	C	A	G	C	T	G	G	T	G	G	G	T	C	G	C	A	T	mmFATP4
1320	C	A	T	C	C	A	G	A	A	G	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T	T	C	G	A	hsFATP4
1320	C	A	T	C	C	A	G	C	A	G	A	C	C	C	T	C	T	G	C	G	C	C	G	T	T	T	C	G	A	mmFATP4	
1350	T	G	G	C	T	A	C	C	T	C	A	A	C	C	A	G	G	G	C	G	C	C	A	A	C	A	A	C	A	A	hsFATP4
1350	C	G	G	G	T	A	C	C	T	C	A	A	C	C	A	G	G	G	T	G	C	C	A	A	C	A	A	C	A	A	mmFATP4
1380	G	A	A	G	A	T	T	G	C	C	A	A	G	A	T	G	T	C	T	T	C	A	A	G	A	A	G	G	G	hsFATP4	
1380	G	A	A	G	A	T	T	G	C	T	A	A	T	G	A	T	G	T	C	T	T	C	A	A	G	A	A	G	G	mmFATP4	
1410	G	G	A	C	C	A	G	C	C	T	A	C	C	T	T	A	C	T	G	G	T	G	A	T	G	T	G	C	T	hsFATP4	
1410	G	G	A	C	C	A	A	G	C	C	T	A	C	C	T	C	A	C	T	G	G	T	G	A	C	G	T	C	T	mmFATP4	
1440	G	G	T	G	A	T	G	G	A	C	G	A	G	C	T	G	G	G	C	T	A	C	C	T	G	T	A	C	T	hsFATP4	
1440	G	G	T	G	A	T	G	G	A	T	G	A	G	C	T	G	G	G	T	T	A	C	C	T	G	T	A	C	T	mmFATP4	

FIG. 31G

1470	C	C	G	A	G	A	C	C	G	A	C	T	G	G	G	G	G	A	C	A	C	G	T	T	C	C	G	C	T	G	hsFATP4	
1470	C	C	G	A	G	A	T	C	G	C	A	C	T	G	G	G	G	A	C	A	C	G	T	T	C	C	G	C	T	G	mmFATP4	
1500	G	A	A	A	G	G	T	G	A	G	A	A	C	G	T	C	C	A	C	C	A	C	C	A	C	G	A	G	G	T	hsFATP4	
1500	G	A	A	A	G	G	T	G	A	G	A	A	T	G	T	A	T	C	T	A	C	C	A	C	T	G	A	G	G	T	mmFATP4	
1530	G	G	A	A	G	G	C	A	C	A	C	T	C	A	G	C	C	G	C	C	C	T	G	C	T	G	G	A	C	A	T	hsFATP4
1530	G	G	A	A	G	G	C	A	C	A	C	T	C	A	G	C	C	G	C	C	T	G	C	T	T	C	A	T	A	T	mmFATP4	
1560	G	G	C	T	G	A	C	G	T	G	G	C	C	G	T	G	T	A	T	G	G	T	G	T	C	G	A	G	G	T	hsFATP4	
1560	G	G	C	A	G	A	T	G	T	G	G	C	A	G	T	T	A	T	G	G	T	G	T	T	G	A	G	G	T	mmFATP4		
1590	G	C	C	A	G	G	A	A	C	C	G	A	G	G	C	C	G	G	C	A	T	G	G	C	C	G	A	A	T	G	C	hsFATP4
1590	G	C	C	A	G	G	A	A	C	T	G	A	A	G	G	C	C	G	A	G	C	A	G	G	A	A	T	G	G	C	mmFATP4	
1620	T	G	C	T	G	T	G	G	C	C	A	G	C	C	C	C	A	C	T	G	G	C	A	A	C	T	G	T	G	A	hsFATP4	
1620	T	G	C	C	G	T	T	G	C	A	A	G	T	C	A	G	C	A	A	C	T	G	T	G	T	G	A	A	A	A	mmFATP4	
1650	C	C	T	G	G	A	G	C	G	C	T	T	T	G	C	T	C	A	G	G	T	C	T	T	G	G	A	A	A	A	hsFATP4	
1650	C	C	T	G	G	A	G	A	G	C	T	T	T	G	C	A	C	A	G	A	C	C	T	T	G	A	A	A	A	A	mmFATP4	

FIG. 31H



1680	GG	A	C	T	G	C	C	C	C	T	G	T	A	T	G	C	G	C	C	C	C	A	T	C	T	T	hsFATP4				
1680	GG	A	G	C	T	G	C	C	T	C	T	G	T	A	T	G	C	C	C	C	C	A	T	C	T	T	mmFATP4				
1710	C	C	T	G	C	G	C	C	T	C	T	G	C	C	T	G	A	G	C	T	G	C	A	A	A	A	C	hsFATP4			
1710	C	C	T	G	C	G	C	T	T	C	T	T	G	C	C	T	G	A	G	C	T	G	C	A	A	A	A	C	mmFATP4		
1740	AG	G	A	A	C	C	T	A	C	A	A	G	T	T	C	C	A	G	A	A	G	A	C	A	G	A	C	T	hsFATP4		
1740	AG	G	A	C	C	T	T	C	A	A	G	T	T	C	C	A	G	A	A	G	A	C	A	G	A	C	T	T	mmFATP4		
1770	A	C	G	G	A	A	G	G	A	G	G	G	C	T	T	G	A	C	C	C	G	G	C	T	A	T	T	G	T	hsFATP4	
1770	G	C	G	G	A	A	G	G	A	G	G	G	C	T	T	G	A	C	C	C	A	T	C	T	G	T	T	G	T	mmFATP4	
1800	G	A	A	G	A	C	C	C	G	C	T	G	T	T	C	T	A	T	C	T	A	G	A	T	G	C	C	A	hsFATP4		
1800	G	A	A	G	A	C	C	C	G	C	T	G	T	T	C	T	A	T	C	T	G	G	A	T	G	C	T	C	G	mmFATP4	
1830	G	A	A	G	G	G	C	C	G	C	T	A	C	G	T	C	C	C	G	A	C	C	A	A	A	G	A	G	A	hsFATP4	
1830	G	A	A	G	G	G	C	T	G	C	T	A	C	G	T	T	G	C	A	C	T	G	G	A	C	C	A	G	A	mmFATP4	
1860	G	G	C	C	T	A	C	A	G	C	C	G	C	A	T	C	C	A	G	G	C	A	G	G	A	G	G	A	G	A	hsFATP4
1860	G	G	C	C	T	A	T	A	C	C	G	C	A	T	C	C	A	G	G	C	A	G	G	C	G	A	G	G	A	mmFATP4	
1890	G	A	A	G	C	T	G																					hsFATP4			
1890	G	A	A	G	C	T	G																					mmFATP4			

FIG. 311

1	M R A P G A G A A S V V S L A L L W L L	hFATP1.
1	M R A P G A G T A S V A S L A L L W F L	mmFATP1
21	G L P W T W S A A A A L G V Y V G S G G	hFATP1.
21	G L P W T W S A A A A F C V Y V G G G G	mmFATP1
41	W R F L R I V C K T A R R D L F G L S V	hFATP1.
41	W R F L R I V C K T A R R D L F G L S V	mmFATP1
61	L I R V R L E L R R H Q R A G H T I P R	hFATP1.
61	L I R V R L E L R R H R R A G D T I P C	mmFATP1
81	I F Q A V V Q R Q P E R L A L V D A G T	hFATP1.
81	I F Q A V A R R Q P E R L A L V D A S S	mmFATP1
101	G E C W T F A Q L D A Y S N A V A N L F	hFATP1.
101	G I C W T F A Q L D T Y S N A V A N L F	mmFATP1
121	R Q L G F A P G D V V A I F L E G R P E	hFATP1.
121	R Q L G F A P G D V V A V F L E G R P E	mmFATP1
141	F V G L W L G L A K A G M E A A L L N V	hFATP1.
141	F V G L W L G L A K A G V V A A L L N V	mmFATP1
161	N L R R E P L A F C L G T S G A K A L I	hFATP1.
161	N L R R E P L A F C L G T S A A K A L I	mmFATP1
181	F G G E M V A A V A E V S G H L G K S L	hFATP1.
181	Y G G E M A A A V A E V S E Q L G K S L	mmFATP1
201	I K F C S G D L G P E G I L P D T H L L	hFATP1.
201	L K F C S G D L G P E S I L P D T Q L L	mmFATP1
221	D P L L K E A S T A P L A Q I P S K G M	hFATP1.
221	D P M L A E A P T T P L A Q A P G K G M	mmFATP1
241	D D R L F Y I Y T S G T T G L P K A A I	hFATP1.
241	D D R L F Y I Y T S G T T G L P K A A I	mmFATP1
261	V V H S R Y Y R M A A F G H H A Y R M Q	hFATP1.
261	V V H S R Y Y R I A A F G H H S Y S M R	mmFATP1

FIG. 32A

281	A A D V L Y D C L P L Y H S A G N I	I G	hFATP1.
281	A A D V L Y D C L P L Y H S A G N I	M G	mmFATP1
301	V G Q C L I Y G L T V V L R K K F S A S		hFATP1.
301	V G Q C V I Y G L T V V L R K K F S A S		mmFATP1
321	R F W D D C I K Y N C T V V Q Y I G E I		hFATP1.
321	R F W D D C V K Y N C T V V Q Y I G E I		mmFATP1
341	C R Y L L K Q P V R E A E R R H R V R L		hFATP1.
341	C R Y L L R Q P V R D V E Q R H R V R L		mmFATP1
361	A V G N G L R P A I W E E F T E R F G V		hFATP1.
361	A V G N G L R P A I W E E F T Q R F G V		mmFATP1
381	R Q I G E F Y G A T E C N C S I A N M D		hFATP1.
381	P Q I G E F Y G A T E C N C S I A N M D		mmFATP1
401	G K V G S C G F N S R I L P H V Y P I R		hFATP1.
401	G K V G S C G F N S R I L T H V Y P I R		mmFATP1
421	L V K V N E D T M E L L R D A Q G L C I		hFATP1.
421	L V K V N E D T M E P L R D S E G L C I		mmFATP1
441	P C Q A G E P G L L V G Q I N Q Q D P L		hFATP1.
441	P C Q P G E P G L L V G Q I N Q Q D P L		mmFATP1
461	R R F D G Y V S E S A T S K K I A H S V		hFATP1.
461	R R F D G Y V S D S A T N K K I A H S V		mmFATP1
481	F S K G D S A Y L S G D V L V M D E L G		hFATP1.
481	F R K G D S A Y L S G D V L V M D E L G		mmFATP1
501	Y M Y F R D R S G D T F R W R G E N V S		hFATP1.
501	Y M Y F R D R S G D T F R W R G E N V S		mmFATP1
521	T T E V E G V L S R L L G Q T D V A V Y		hFATP1.
521	T T E V E A V L S R L L G Q T D V A V Y		mmFATP1
541	G V A V P G V E G K A G M A A V A D P H		hFATP1.
541	G V A V P G V E G K A G M A A I A D P H		mmFATP1

FIG. 32B

561	S	L	L	D	P	N	A	I	Y	Q	E	L	Q	K	V	L	A	P	Y	A	hFATP1.
561	S	Q	L	D	P	N	S	M	Y	Q	E	L	Q	K	V	L	A	S	Y	A	mmFATP1
581	R	P	I	F	L	R	L	L	P	Q	V	D	T	T	G	T	F	K	I	Q	hFATP1.
581	R	P	I	F	L	R	L	L	P	Q	V	D	T	T	G	T	F	K	I	Q	mmFATP1
601	K	T	R	L	Q	R	E	G	F	D	P	R	Q	T	S	D	R	L	F	F	hFATP1.
601	K	T	R	L	Q	R	E	G	F	D	P	R	Q	T	S	D	R	L	F	F	mmFATP1
621	L	D	L	K	Q	G	H	Y	L	P	L	N	E	A	V	Y	T	R	I	C	hFATP1.
621	L	D	L	K	O	G	R	Y	V	P	L	D	E	R	V	H	A	R	I	C	mmFATP1
641	S	G	A	F	A	L														hFATP1.	
641	A	G	D	F	S	L														mmFATP1	

FIG. 32C

L	F	S	K	L	V	L	K	L	P	W	T	Q	V	G	F	S	L	L	hsFATP4pep
L	G	S	K	L	V	L	K	L	P	W	T	Q	V	G	F	S	L	L	nmFATP4pep
F	L	Y	L	G	S	G	G	W	R	F	I	R	V	F	I	K	T	I	hsFATP4pep
L	L	Y	L	G	S	G	G	W	R	F	I	R	V	F	I	K	T	V	nmFATP4pep
K	V	K	A	K	V	R	Q	C	L	Q	E	R	R	T	V	P	I	L	hsFATP4pep
K	V	K	T	K	V	R	R	Y	L	Q	E	R	K	T	V	P	L	L	nmFATP4pep
T	A	L	I	F	E	G	T	D	T	H	W	T	F	R	Q	L	D	E	hsFATP4pep
T	A	L	I	F	E	G	T	D	T	H	W	T	F	R	Q	L	D	E	nmFATP4pep
R	G	L	A	S	G	D	V	A	I	F	M	E	N	R	N	E	F	V	hsFATP4pep
R	G	L	A	S	G	N	V	V	A	L	F	M	E	N	R	N	E	F	nmFATP4pep
V	E	A	A	L	I	N	T	N	L	R	R	D	A	L	L	H	C	L	hsFATP4pep
V	E	A	A	L	I	N	T	N	L	R	R	D	A	L	L	R	H	C	nmFATP4pep
S	E	M	A	S	A	I	C	E	V	H	A	S	L	D	P	S	L	S	hsFATP4pep
S	E	M	A	S	A	I	C	E	I	H	A	S	L	E	P	T	L	S	nmFATP4pep

FIG. 33A

V P	P S T E H L D D P L L K	D A P K H L P S	C P D K G F T D K	hsFATP4pep
V P	V S T E H L D D P L L E	D A P K H L P S	H P D K G F T D K	mmFATP4pep
L F Y I Y T S G T T G L P K A A I V V H S R Y Y R M A A L V	L F Y I Y T S G T T G L P K A A I V V H S R Y Y R M A A S L V			hsFATP4pep
				mmFATP4pep
Y Y G F R M R P N D I V Y D C L P L Y H S A G N I V G I G Q	Y Y G F R M R P D D I V Y D C L P L Y H S S R K H R G D W Q			hsFATP4pep
				mmFATP4pep
C L L H G M T V V I R K K F S A S R F W D D C I K Y N C T I	C L L H G M T V V I R K K F S A S R F W D D C I K Y N C T V			hsFATP4pep
				mmFATP4pep
V Q Y I G E L C R Y L L N Q P P R E A E N Q H Q	V Q Y I G E L C R Y L L N Q P P R E A E S R H K			hsFATP4pep
				mmFATP4pep
N G L R Q S I W T N F S S R F H I P Q V A E F Y G A T E C N	N G L R Q S I W T D F S S R F H I P Q V A E F Y G A T E C N			hsFATP4pep
				mmFATP4pep
C S L G N F D S Q V G A C G F N S R I L S F V Y P I R L V R	C S L G N F D S R V G A C G F N S R I L S F V Y P I R L V R			hsFATP4pep
				mmFATP4pep

FIG. 33B

V N E D T M E L I R G P D G V C I P C Q P G	E P G Q L V G R	hsFATP4pep
V N E D T M E L I R G P D G V C I P C Q P G	Q P G Q L V G R	mmFATP4pep
I I Q K D P L R R F D G Y L N Q G A N N K K I A	K D V F K K	hsFATP4pep
I I Q Q D P L R R F D G Y L N Q G A N N K K I A	N D V F K K	mmFATP4pep
G D Q A Y L T G D V L V M D E L G Y L Y F R D R T G D T F R		hsFATP4pep
G D Q A Y L T G D V L V M D E L G Y L Y F R D R T G D T F R		mmFATP4pep
W K G E N V S T T E V E G T L S R L L D	M A D V A V Y G V E	hsFATP4pep
W K G E N V S T T E V E G T L S R L L H	M A D V A V Y G V E	mmFATP4pep
V P G T E G R A G M A A V A S P	T G N C D L E R F A Q V L E	hsFATP4pep
V P G T E G R A G M A A V A S P	I S N C D L E S F A Q T L K	mmFATP4pep
K E L P L Y A R P I F L R	L L P E L H K T G T Y K F Q K T E	hsFATP4pep
K E L P L Y A R P I F L R	F L P E L H K T G T F K F Q K T E	mmFATP4pep
L R K E G F D P A I V K D P L F Y L D A Q	K G R Y V P L D Q	hsFATP4pep
L R K E G F D P S V V K D P L F Y L D A R	K G C Y V A L D Q	mmFATP4pep
E A Y S R I Q A G E E K L		hsFATP4pep
E A Y T R I Q A G E E K L		mmFATP4pep

FIG. 33C

```

1   aac ggc aag taa gcg caa cgc aat taa tgt gag tag ctc act cat tag gca ccc cag gct
61  tta cac ttt atg ctt ccg ggc tcg tat gtt gtg tgg aat tgt gag cgg ata cca att tca
121 cac agg aac cag cta tga cat gat tac gaa ttt aat acg act cac tat agg gaa ttt ggc
181 cct cga ggc caa gaa ttc ggc acg agg ggt gct gag ccc ctg cgc ggt ttc tgg tgc gta
241 gag act gta aat cgc tgc gct tct cag tca tca tcc cag ctt ttc ccg gct cga att
301 cag cct cca act caa gct cgc ggc aaa gac tac ctg aga gga gaa aag ctt ctg tcc ctg
361 gac ctt ctt ctg agg gtg gag tcg gag gct ccc tgc ttt cca gcc cag tga ccc aag
421 ctt aat ctt cag cac cac ttg ggc cga cct ttt cgg tgc aaa cct acg att ctg ttt ctc
481 agg att cct ccc cat ccc gct tcg ccc cgg aaa agc tga caa gaa ctt cag gtg taa gcc
541 ctg agt agt gag gat ctg cgg tct ccg tgg aga gct gtg cct gga aga gaa gga cgc tgg
601 tgg ggc ctg aga tca gag ctg tct tct ggc cca gtt gcc ccc atg ctt ctg tca tgg cta

                                     M L L S W L
661  aca gtt cta ggg gct gga atg gtc gtc ctg cac ttc ttg cag aaa ctc ctg ttc cct tac
    T V L G A G M V V L H F L Q K L L F P Y
721  ttt tgg gat gac ttc tgg ttc gtg ttg aag gtg gtg ctc att ata att cgg ctg aag aag
    F W D D F W F V L K V V L I I I R L K K
781  tat gaa aag aga ggg gag ctg gtg act gtg ctg gat aaa ttc ttg agt cat gcc aaa aga
    Y E K R G E L V T V L D K F L S H A K R

```

FIG. 34A



841 caa cct cgg aaa cct ttc atc atc tat gag gga gac atc tac acc tat cag gat gta gac  
 Q P R K P F I I Y E G D I Y T Y Q D V D  
 901 aaa agg agc agc aga gtg gcc cat gtc ttc ctg aac cat tcc tct ctg aaa aag ggg gac  
 K R S S R V A H V F L N H S S L K K G D  
 961 acg gtg gct ctg ctg atg agc aat gag ccg gac ttc gtt cac gtg tgg ttc ggc ctg gcc  
 T V A L L M S N E P D F V H V W F G L A  
 1021 aag ctg ggc tgc gtg gtg gcc ttt ctg aac acc aac att cgc tcc aac tcc ctg ctg aat  
 K L G C V V A F L N T N I R S N S L L N  
 1081 tgc atc cgc gcc tgt ggt gcc aga gcc cta gtg gtg ggc gca gat ttg ctt gga acg gta  
 C I R A C G P R A L V V G A D L L G T V  
 1141 gaa gaa atc ctt cca agc ctc tca gaa aat atc agt gtt tgg ggg atg aaa gat tct gtt  
 E E I L P S L S E N I S V W G M K D S V  
 1201 cca caa ggt gta att tca ctc aaa gaa aaa ctg agc acc tca cct gat gag ccc gtg cca  
 P Q G V I S L K E K L S T S P D E P V P  
 1261 cgc agc cac cat gtt gtc tca ctc ctg aag tct act tgt ctt tac att ttt acc tct gg  
 R S H H V V S L L K S T C L Y I F T S G  
 1321 aca aca ggt cta cca aaa gca gct gtg att agt cag ctg cag gtt tta agg ggt tct gc  
 T T G L P K A A V I S Q L Q V L R G S A  
 1381 gtc ctg tgg gct ttt ggt tgt act gct cat gac att gtt tat ata acc ctt cct ctg ta  
 V L W A F G C T A H D I V Y I T L P L Y

FIG. 34B

1441 cat agt tca gca gct atc ctg gga att tct gga tgt. gtt gag ttg ggt gcc act tgt gtg  
 H S A A I L G I S G C V E L G A T C V  
 1501 tta aag aag aaa ttt tca gca agc cag ttt tgg agt gac tgc aag aag tat gat gtg act.  
 L K K F S A S Q F W S D C K K Y D V T  
 1561 gtg ttt cag tat att gga gaa ctt tgt cgc tac ctt tgc aaa caa tct aag aga gaa gga  
 V F Q Y I G E L C R Y L C K Q S K R E G  
 1621 gaa aag gat cat aag gtg cgt ttg gca att gga aat ggc ata. cgg agt gat gta tgg aga  
 E K D H K V R L A I G N G I R S D V W R  
 1681 gaa ttt tta gac aga ttt gga aat ata aag gtg tgt gaa ctt tat gca gct acc gaa tca  
 E F L D R F G N I K V C E L Y A A T E S  
 1741 agc ata tct ttc atg aac tac act ggg aga att gga gca att ggg aga aca aat ttg ttt  
 S I S F M N Y T G R I G A I G R T N L F  
 1801 tac aaa ctt ctt tcc act ttt gac tta ata aag tat gac ttt cag aaa gat gaa ccc atg  
 Y K L L S T F D L I K Y D F Q K D E P M  
 1861 aga aat gag cag ggt tgg tgt att cat gtg aaa aag gga gaa cct gga ctt ctc att tct  
 R N E Q G W C I H V K K G E P G L L I S  
 1921 cga gtg aat gca aaa aat ccc ttc ttt ggc tat gct ggg cct tat aag cac aca aaa gac  
 R V N A K N P F F G Y A G P Y K H T K D  
 1981 aaa ttg ctt tgt gat gtt ttt aag aag gga gat gtt tac. ctt aat act gga gac tta ata  
 K L L C D V F K K G D V Y L N T G D L I  
 2041 gtc cag gat cag gac aat ttc ctt tat ttt tgg gac cgt act gga gac act ttc aga tgg  
 V Q D Q D N F L Y F W D R T G D T F R W  
 2101 aaa gga gaa aat gtc gca acc act gag gtt gct gat gtt att gga atg ttg gat ttc ata

FIG. 34C

2161 K G E N V A T T E V A D V I G M L D F I  
 cag gaa gca aac gtc tat ggt gtg gct ata tca ggt tat gaa gga aga gca gga atg gct  
 Q E A N V Y G V A I S G Y E G R A G M A  
 2221 tct att att tta aaa cca aat aca tct tta gat ttg gaa aaa gtt tat gaa caa gtt gta  
 S I I L K P N T S L D L E K V Y E Q V V  
 2281 aca ttt cta cca gct tat gct tgt cca cga ttt tta aga att cag gaa aaa atg gaa gca  
 T F L P A Y A C P R F L R I Q E K M E A  
 2341 aca gga aca ttc aaa cta ttg aag cat cag ttg gtg gaa gat gga ttt aat cca ctg aaa  
 T G T F K L L K H Q L V E D G F N P L K  
 2401 att tct gaa cca ctt tac ttc atg gat aac ttg aaa aag tct tat gtt cta ctg acc agg  
 I S E P L Y F M D N L K K S Y V L L T R  
 2461 gaa ctt tat gat caa ata atg tta ggg gaa ata aaa ctt taa gat ttt tat atc tag aac  
 E L Y D Q I M L G E I K L \*  
 2521 ttt cat atg ctt tct tag gaa gag tga gag ggg ggt ata tga ttc ttt atg aaa tgg gga  
 2581 aag gga gct aac att aat tat gca tgt act ata ttt cct taa tat gag aga taa ttt ttt  
 2641 aat tgc ata aga att tta att tct ttt aat tga tat aaa cat tag ttg att att ctt ttt  
 2701 atc tat ttg gag att cag tgc ata act aag tat ttt cct taa tac taa aga ttt taa ata  
 2761 ata aat agt ggc tag cgg ttt gga caa tca cta aaa atg tac ttt cta ata agt aaa att  
 2821 tct aat ttt gaa taa aag att aaa ttt tac tga aaa aaa aaa atg aaa atg ggc  
 2881 gcc gc

FIG. 34D

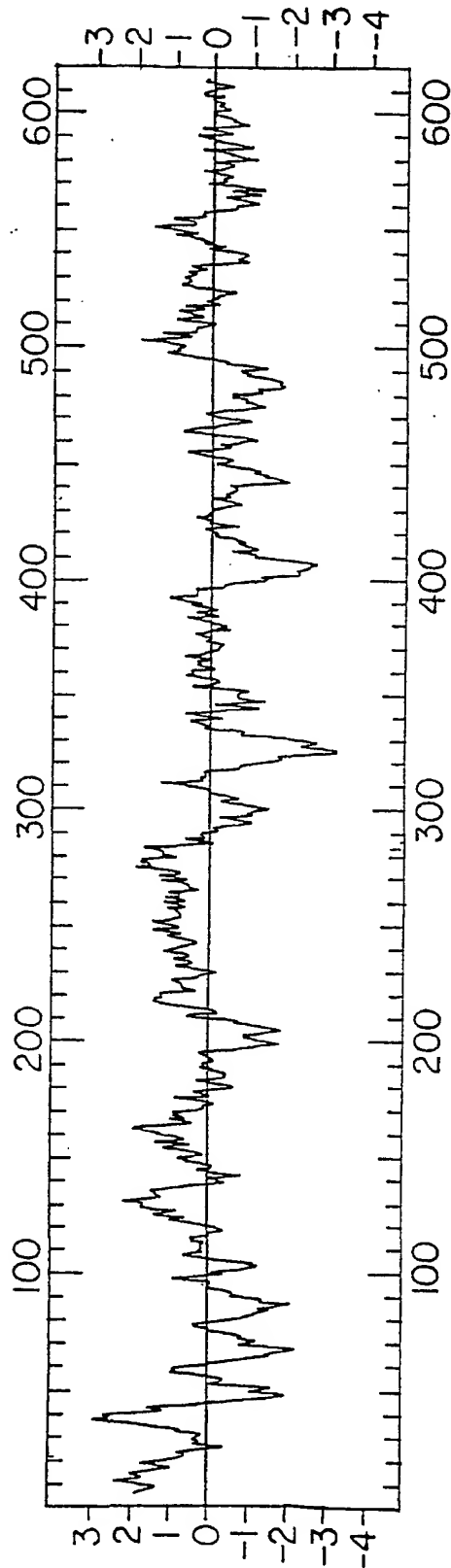


FIG. 35A

		<u>n</u>	<u>n%</u>	<u>MW</u>	<u>MW%</u>
A ala	alanine	33	5.3	2344	3.3
C cys	cysteine	14	2.3	1442	2.1
D asp	aspartic acid	34	5.5	3910	5.6
E glu	glutamic acid	31	5.0	4000	5.7
F phe	phenylalanine	34	5.5	5000	7.1
G gly	glycine	44	7.1	2508	3.6
H his	histidine	13	2.1	1781	2.5
I ile	isoleucine	37	6.0	4184	6.0
K lys	lysine	48	7.8	6148	8.8
L leu	leucine	75	12.1	8481	12.1
M met	methionine	11	1.8	1441	2.1
N asn	asparagine	21	3.4	2394	3.4
P pro	proline	21	3.4	2038	2.9
Q gln	glutamine	18	2.9	2305	3.3
R arg	arginine	27	4.4	4214	6.0
S ser	serine	40	6.5	3481	5.0
T thr	threonine	30	4.8	3031	4.3
V val	valine	51	8.2	5052	7.2
W trp	tryptophan	11	1.8	2046	2.9
X ukw	unknown	--	--		
Y tyr	tyrosine	26	4.2	4239	6.1
Z ---	STOP				

FIG. 35B

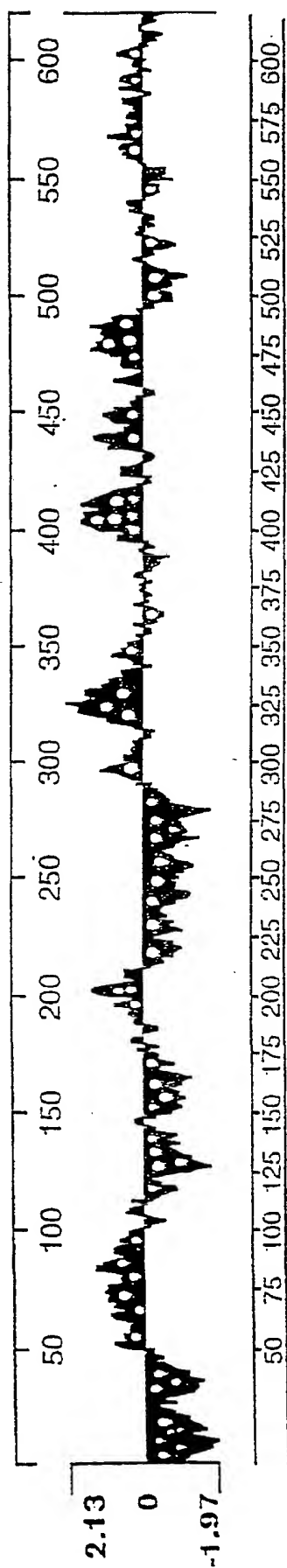


FIG. 35C

1	M	R	A	P	-	-	G	A	G	A	A	S	V	S	L	A	L	L	W	hsFATP1pep		
1	L	-	-	-	-	-	-	-	-	-	-	-	-	-	F	S	K	L	-	hsFATP4pep		
1	M	L	L	S	W	L	T	V	L	G	A	G	M	V	V	L	H	F	L	Q	hsFATP6pep	
19	L	L	G	L	P	W	T	W	S	A	A	A	A	L	G	V	Y	V	G	S	hsFATP1pep	
6	V	L	K	L	P	W	T	Q	V	G	F	S	L	L	F	L	Y	L	G	S	hsFATP4pep	
21	K	L	L	F	P	Y	F	W	D	D	F	-	-	-	-	-	-	-	-	-	hsFATP6pep	
39	G	G	W	R	F	L	R	I	V	C	K	T	A	R	R	D	L	F	G	L	hsFATP1pep	
26	G	G	W	R	F	I	R	V	F	I	K	T	I	R	R	D	I	F	G	G	hsFATP4pep	
32	-	-	W	F	V	L	K	V	-	-	-	-	-	-	-	-	-	-	-	-	hsFATP6pep	
59	S	V	L	I	R	V	R	L	E	L	R	R	H	Q	R	A	G	H	T	I	hsFATP1pep	
46	L	V	L	L	K	V	K	A	K	V	R	Q	C	L	Q	E	R	R	T	V	hsFATP4pep	
38	-	V	L	I	I	I	R	L	K	K	Y	E	K	R	G	E	L	V	T	V	hsFATP6pep	
79	P	R	I	F	Q	A	V	V	Q	R	Q	P	E	R	L	A	L	V	D	A	hsFATP1pep	
66	P	I	L	F	A	S	T	V	R	R	H	P	P	D	K	T	A	L	I	F	E	hsFATP4pep
57	L	D	K	F	L	S	H	A	K	R	Q	P	P	R	K	P	F	I	I	Y	E	hsFATP6pep

FIG. 36A

99	G	T	G	E	C	W	T	F	A	Q	L	D	A	Y	S	N	A	V	A	N	hsFATP1pep
86	G	T	D	T	H	W	T	F	R	Q	L	D	E	Y	S	S	S	V	A	N	hsFATP4pep
77	G	-	-	D	I	Y	T	Y	Q	D	V	D	K	R	S	S	R	V	A	H	hsFATP6pep
119	-	L	F	R	Q	L	G	F	A	P	G	D	V	V	A	I	F	L	E	G	hsFATP1pep
106	-	F	L	Q	A	R	G	L	A	S	G	D	V	A	A	I	F	M	E	N	hsFATP4pep
95	V	F	L	N	H	S	S	L	K	K	G	D	T	V	A	L	L	M	S	N	hsFATP6pep
138	R	P	E	F	V	G	L	W	L	G	L	A	K	A	G	M	E	A	A	L	hsFATP1pep
125	R	N	E	F	V	G	L	W	L	G	M	A	K	L	G	V	E	A	A	L	hsFATP4pep
115	E	P	D	E	V	H	V	W	F	G	L	A	K	L	G	C	V	V	A	F	hsFATP6pep
158	L	N	V	N	L	R	R	E	P	L	A	F	C	L	G	T	S	G	A	K	hsFATP1pep
145	I	N	T	N	L	R	R	D	A	L	L	H	C	L	T	T	S	R	A	R	hsFATP4pep
135	L	N	T	N	I	R	S	N	S	L	L	N	C	I	R	A	C	G	P	R	hsFATP6pep
178	A	L	I	F	G	G	E	M	V	A	A	V	A	E	V	S	G	H	L	G	hsFATP1pep
165	A	L	V	F	G	S	E	M	A	S	A	I	C	E	V	H	A	S	L	D	hsFATP4pep
155	A	L	V	V	G	A	D	L	L	G	T	V	E	E	I	L	P	S	L	S	hsFATP6pep

FIG. 36B



198	K	S	L	I	K	F	C	S	G	D	L	G	P	E	G	I	L	P	D	T	hsFATP1pep
185	P	S	L	S	L	F	C	S	G	S	W	E	P	G	A	V	P	P	S	T	hsFATP4pep
175	E	N	I	S	V	W	G	M	K	D	S	V	P	Q	G	V	I	S	-	-	hsFATP6pep
218	H	L	L	D	P	L	L	K	E	A	S	T	A	P	L	A	Q	I	P	S	hsFATP1pep
205	E	H	L	D	P	L	L	K	D	A	P	K	-	H	L	P	S	C	P	D	hsFATP4pep
193	-	-	L	K	E	K	L	S	T	S	P	D	E	P	V	P	R	S	H	H	hsFATP6pep
238	K	G	-	-	M	D	D	R	L	F	Y	I	Y	T	S	G	T	T	G	L	hsFATP1pep
224	K	G	-	-	F	T	D	K	L	F	Y	I	Y	T	S	G	T	T	G	L	hsFATP4pep
211	V	V	S	L	L	K	S	T	C	L	Y	I	F	T	S	G	T	T	G	L	hsFATP6pep
256	P	K	A	A	I	V	V	H	S	R	Y	Y	R	M	A	A	F	G	H	H	hsFATP1pep
242	P	K	A	A	I	V	V	H	S	R	Y	Y	R	M	A	A	L	V	Y	Y	hsFATP4pep
231	P	K	A	A	V	I	S	Q	L	Q	V	L	R	G	S	A	-	V	L	W	hsFATP6pep
276	A	Y	R	M	Q	A	A	D	V	L	Y	D	C	L	P	L	Y	H	S	A	hsFATP1pep
262	G	F	R	M	R	P	N	D	I	V	Y	D	C	L	P	L	Y	H	S	A	hsFATP4pep
250	A	F	G	C	T	A	H	D	I	V	Y	I	T	L	P	L	Y	H	S	S	hsFATP6pep

FIG. 36C

296	G	N	I	I	G	V	G	Q	C	L	I	Y	G	L	T	V	V	L	R	K	hsFATP1pep
282	G	N	I	V	G	I	G	Q	C	L	L	H	G	M	T	V	V	I	R	K	hsFATP4pep
270	A	A	I	L	G	I	S	G	C	V	E	L	G	A	T	C	V	L	K	K	hsFATP6pep
316	K	F	S	A	S	R	F	W	D	D	C	I	K	Y	N	C	T	V	V	Q	hsFATP1pep
302	K	F	S	A	S	R	F	W	D	D	C	I	K	Y	N	C	T	I	V	Q	hsFATP4pep
290	K	F	S	A	S	Q	F	W	S	D	C	K	K	Y	D	V	T	V	F	Q	hsFATP6pep
336	Y	I	G	E	I	C	R	Y	L	L	K	Q	P	V	R	E	A	E	R	R	hsFATP1pep
322	Y	I	G	E	L	C	R	Y	L	L	N	Q	P	P	R	E	A	E	N	Q	hsFATP4pep
310	Y	I	G	E	L	C	R	Y	L	C	K	Q	S	K	R	E	G	E	K	D	hsFATP6pep
356	H	R	V	R	L	A	V	G	N	G	L	R	P	A	I	W	E	E	F	T	hsFATP1pep
342	H	Q	V	R	M	A	L	G	N	G	L	R	Q	S	I	W	T	N	F	S	hsFATP4pep
330	H	K	V	R	L	A	I	G	N	G	I	R	S	D	V	W	R	E	F	L	hsFATP6pep
376	E	R	F	G	V	R	Q	I	G	E	F	Y	G	A	T	E	C	N	C	S	hsFATP1pep
362	S	R	F	H	I	P	Q	V	A	E	F	Y	G	A	T	E	C	N	C	S	hsFATP4pep
350	D	R	F	G	N	I	K	V	C	E	L	Y	A	A	T	E	S	S	I	S	hsFATP6pep

FIG. 36D

396	I	A	N	M	D	G	K	V	G	S	C	G	F	N	S	R	I	L	P	H	hsFATP1pep
382	L	G	N	F	D	S	Q	V	G	A	C	G	F	N	S	R	I	L	S	F	hsFATP4pep
370	F	M	N	Y	T	G	R	I	G	A	I	G	R	T	N	L	F	Y	K	L	hsFATP6pep
416	V	Y	P	I	R	L	V	K	V	N	E	D	T	M	E	L	L	R	D	A	hsFATP1pep
402	V	Y	P	I	R	L	V	R	V	N	E	D	T	M	E	L	I	R	G	P	hsFATP4pep
390	L	S	T	F	D	L	I	K	Y	D	F	Q	K	D	E	P	M	R	N	E	hsFATP6pep
436	Q	G	L	C	T	P	C	Q	A	G	E	P	G	L	L	V	G	Q	I	N	hsFATP1pep
422	D	G	V	C	I	P	C	Q	P	G	E	P	G	Q	L	V	G	R	I	I	hsFATP4pep
410	Q	G	W	C	I	H	V	K	K	G	E	P	G	L	L	I	S	R	V	N	hsFATP6pep
456	Q	Q	D	P	L	R	R	F	D	G	Y	V	S	E	S	A	T	S	K	-	hsFATP1pep
442	Q	K	D	P	L	R	R	F	D	G	Y	L	N	Q	G	A	N	N	K	-	hsFATP4pep
430	A	K	N	P	-	-	-	F	F	G	Y	A	G	P	Y	K	H	T	K	D	hsFATP6pep
475	K	I	A	H	S	V	F	S	K	G	D	S	A	Y	L	-	S	G	D	V	hsFATP1pep
461	K	I	A	K	D	V	F	K	K	G	D	Q	A	Y	L	-	T	G	D	V	hsFATP4pep
447	K	I	L	C	D	V	F	K	K	G	D	-	V	Y	L	N	T	G	D	L	hsFATP6pep

FIG. 36E

494	L	V	M	D	E	L	G	Y	M	Y	F	R	D	R	S	G	D	T	F	R	hsFATP1pep
480	L	V	M	D	E	L	G	Y	L	Y	F	R	D	R	T	G	D	T	F	R	hsFATP4pep
466	I	V	Q	D	Q	D	N	F	L	Y	F	W	D	R	T	G	D	T	F	R	hsFATP6pep
514	W	R	G	E	N	V	S	T	T	E	V	E	G	V	L	S	R	L	L	G	hsFATP1pep
500	W	K	G	E	N	V	S	T	T	E	V	E	G	T	L	S	R	L	L	D	hsFATP4pep
486	W	K	G	E	N	V	A	T	T	E	V	A	D	V	I	G	M	L	D	F	hsFATP6pep
534	Q	T	D	V	A	V	Y	G	V	A	V	P	G	V	E	G	K	A	G	M	hsFATP1pep
520	M	A	D	V	A	V	Y	G	V	E	V	P	G	T	E	G	R	A	G	M	hsFATP4pep
506	I	Q	E	A	N	V	Y	G	V	A	I	S	G	Y	E	G	R	A	G	M	hsFATP6pep
554	A	A	V	A	-	D	P	H	S	L	L	D	P	N	A	I	Y	Q	E	L	hsFATP1pep
540	A	A	V	A	-	S	P	T	G	N	C	D	L	E	R	F	A	Q	V	L	hsFATP4pep
526	A	S	I	I	L	K	P	N	T	S	L	D	L	E	K	V	Y	E	Q	V	hsFATP6pep
573	Q	K	V	L	A	P	Y	A	R	P	I	F	L	R	L	L	P	Q	V	D	hsFATP1pep
559	E	K	E	L	P	L	Y	A	R	P	I	F	L	R	L	L	P	E	L	H	hsFATP4pep
546	V	T	F	L	P	A	Y	A	C	P	R	F	L	R	I	Q	E	K	M	E	hsFATP6pep

FIG. 36F

593	T	T	G	T	F	K	I	Q	K	T	R	L	Q	R	E	G	F	D	P	R	hsFATP1pep
579	K	T	G	T	Y	K	F	Q	K	T	E	L	R	K	E	G	F	D	P	A	hsFATP4pep
566	A	T	G	T	F	K	L	L	K	H	Q	L	V	E	D	G	F	N	P	L	hsFATP6pep
613	Q	T	S	D	R	L	F	F	L	D	L	K	Q	G	H	Y	L	P	L	N	hsFATP1pep
599	I	V	K	D	P	L	F	Y	L	D	A	Q	K	G	R	Y	V	P	L	D	hsFATP4pep
586	K	I	S	E	P	L	Y	F	M	D	N	L	K	K	S	Y	V	L	L	T	hsFATP6pep
633	E	A	V	Y	T	R	I	C	S	G	A	F	A	L							hsFATP1pep
619	Q	E	A	Y	S	R	I	Q	A	G	E	E	K	L							hsFATP4pep
606	R	E	L	Y	D	Q	I	M	L	G	E	I	K	L							hsFATP6pep

FIG. 36G

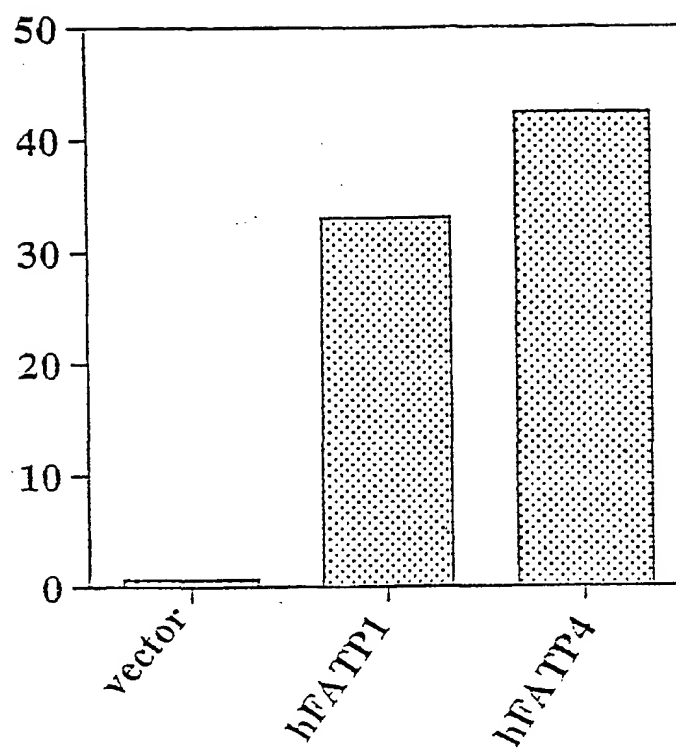


FIG. 37

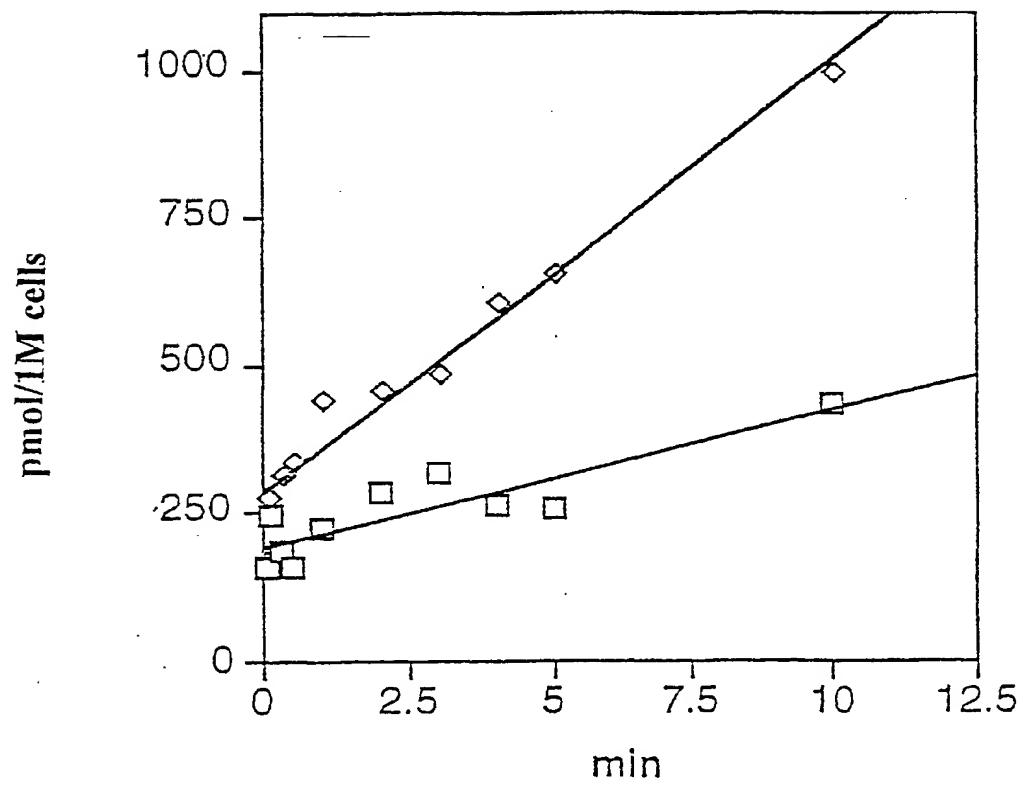


FIG. 38

hsFATP4_1	MLL	-GASLVGVLLFSKL	-VLKLPWTQVGFSL	L	F	31
mmFATP4_1	MLL	-GASLVGVLLFSKL	-VLKLPWTQVGFSL	L	X	31
hsFATP1_1	MRAPGA	GASVVSLL	WLLGLPWTWSAAAL	L	G	33
hsFATP4_32	LYLGSGGWR	FIRVF	IKTIRRD	IFGG	L	64
mmFATP4_32	LYLGSGGWR	FIRVF	IKTVRRD	IFGG	M	64
hsFATP1_34	VYVVGSGGWR	FLRTVC	KTARRDL	FGLSVL	IRVRL	66
hsFATP4_65	KVRQCLQER	RTVPIL	FAS	TVRRHPDK	TALIFEG	97
mmFATP4_65	KVRRLYLQER	KTVP	LLFAS	MVQRHPDK	TALIFEG	97
hsFATP1_67	ELRRRHQRAGH	ITIPRI	FQAVVQRQ	PERLAL	VDA	99
hsFATP4_98	TDTHWTFRQL	DEYSSVANFL	QARGLASGD	V	A	130
mmFATP4_98	TDTHWTFRQL	DEYSSVANFL	QARGLASG	N	VVA	130
hsFATP1_100	TGECWTF	AQLDAYS	NANLFRQL	GFA	PGDVVA	132
hsFATP4_131	IFMENRNEF	VGLWLGMAK	LGVEAALINTNL	RRD		163
mmFATP4_131	LFMENRNEF	VGLWLGMAK	LGVEAALINTNL	RRD		163
hsFATP1_133	IFLEGRPEF	VGLWLG	LAKAGMEAAAL	LVN	LRRE	165
hsFATP4_164	ALHCL	TTSRARAL	VFGSEMASA	ICEVHAS	LDP	196
mmFATP4_164	ALRHCL	DTSKARAL	IFGSEMASA	ICEI	HASLEP	196
hsFATP1_166	PLAFCL	GTSGAKAL	IFGGE	MAVA	EVSGHLGK	198
hsFATP4_197	SLSLFC	SGSWEP	GA	VPPSTEHL	DP	228
mmFATP4_197	TLSLFC	SGSWEP	STVP	VSTEHLD	PL	228
hsFATP1_199	SLIKFC	SGDLGP	EGILP	DTHL	LDPL	231

FIG. 39A



hsFATP4_229	L	P	S	C	P	D	K	G	F	T	D	K	L	F	Y	I	Y	T	S	G	T	T	G	L	P	K	A	A	I	V	V	H	S	261
mmFATP4_229	L	P	S	H	P	D	K	G	F	T	D	K	L	F	Y	I	Y	T	S	G	T	T	G	L	P	K	A	A	I	V	V	H	S	261
hsFATP1_232	L	A	Q	I	P	S	K	G	M	D	D	R	L	F	Y	I	Y	T	S	G	T	T	G	L	P	K	A	A	I	V	V	H	S	264
-----																																		
hsFATP4_262	R	Y	Y	R	M	A	A	L	V	Y	Y	G	F	R	M	R	P	N	D	I	V	Y	D	C	L	P	L	Y	H	S	A	G	N	294
mmFATP4_262	R	Y	Y	R	M	A	S	L	V	Y	Y	G	F	R	M	R	P	D	I	V	Y	D	C	L	P	L	Y	H	S	S	R	K	294	
hsFATP1_265	R	Y	Y	R	M	A	A	F	G	H	A	Y	R	M	Q	A	A	D	V	L	Y	D	C	L	P	L	Y	H	S	A	G	N	297	
-----																																		
hsFATP4_295	I	V	G	I	G	Q	C	L	L	H	G	M	T	V	I	R	K	K	F	S	A	S	R	F	W	D	D	C	I	K	Y	N	327	
mmFATP4_295	H	R	G	D	W	Q	C	L	L	H	G	M	T	V	I	R	K	K	F	S	A	S	R	F	W	D	D	C	I	K	Y	N	327	
hsFATP1_298	I	I	G	V	G	Q	C	L	I	Y	G	L	T	V	L	R	K	K	F	S	A	S	R	F	W	D	D	C	I	K	Y	N	330	
-----																																		
hsFATP4_328	C	T	I	V	Q	Y	I	G	E	L	C	R	Y	L	L	N	Q	P	P	R	E	A	E	N	Q	H	Q	V	R	M	A	L	G	360
mmFATP4_328	C	T	V	V	Q	Y	I	G	E	L	C	R	Y	L	L	N	Q	P	P	R	E	A	E	S	R	H	K	V	R	M	A	L	G	360
hsFATP1_331	C	T	V	V	Q	Y	I	G	E	I	C	R	Y	L	L	K	Q	P	V	R	E	A	E	R	R	H	R	V	R	L	A	V	G	363
-----																																		
hsFATP4_361	N	G	L	R	Q	S	I	W	T	N	F	S	S	R	F	H	I	P	Q	V	A	E	F	Y	G	A	T	E	C	N	C	S	L	393
mmFATP4_361	N	G	L	R	Q	S	I	W	T	D	F	S	S	R	F	H	I	P	Q	V	A	E	F	Y	G	A	T	E	C	N	C	S	L	393
hsFATP1_364	N	G	L	R	P	A	I	W	E	E	F	T	E	R	F	G	V	R	Q	I	G	E	F	Y	G	A	T	E	C	N	C	S	I	396
-----																																		
hsFATP4_394	G	N	F	D	S	Q	V	G	A	C	G	F	N	S	R	I	L	S	F	V	Y	P	I	R	L	V	R	V	N	E	D	T	M	426
mmFATP4_394	G	N	F	D	S	R	V	G	A	C	G	F	N	S	R	I	L	S	F	V	Y	P	I	R	L	V	R	V	N	E	D	T	M	426
hsFATP1_397	A	N	M	D	G	K	V	G	S	C	G	F	N	S	R	I	L	P	H	V	Y	P	I	R	L	V	K	V	N	E	D	T	M	429

FIG. 39B

hsFATP4_427	ELIRGPDGVCIP	CQPGEP	GGQLVGR	IIQK	DPPLRR	459				
mmFATP4_427	ELIRGPDGVCIP	CQPGQ	PGQLVGR	IIQQ	DPPLRR	459				
hsFATP1_430	ELLRDAQGLCIP	CQA	GEPLLVG	QIN	QQDPPLRR	462				
hsFATP4_460	FDGYLNQGAN	KKIAK	DVFKKGD	QAYLT	GDVLV	492				
mmFATP4_460	FDGYLNQGAN	KKIAAN	DVFKKGD	QAYLT	GDVLV	492				
hsFATP1_463	FDGYVSES	ATSKKIA	HSVFS	KGDS	AYLSGDVLV	495				
hsFATP4_493	MDELGYLYFR	DRTGDT	FRWK	GENVST	TEVEGTL	525				
mmFATP4_493	MDELGYLYFR	DRTGDT	FRWK	GENVST	TEVEGTL	525				
hsFATP1_496	MDELGYMYFR	DRSGD	TFRWR	GENVST	TEVEGVL	528				
hsFATP4_526	SRLLDMAD	VAVYGV	VEVPG	TEGR	AGMAAVASPTG	558				
mmFATP4_526	SRLLHMA	DVAVYGV	VEVPG	TEGR	AGMAAVASPIIS	558				
hsFATP1_529	SRLLGQ	TDVAVYGV	AVPG	V	EGKAGMAAVADPHS	561				
hsFATP4_559	NCDLER	FAQVL	EKEL	PLYAR	PIFLRLPELHKT	591				
mmFATP4_559	NCDLES	FAQTL	EKEL	PLYAR	PIFLRLPELHKT	591				
hsFATP1_562	LLDP	NAIYQ	ELQK	VLAPYAR	PIFLRLPQVDTT	594				
hsFATP4_592	GT	YK	FQKTEL	RKEG	FDPAIVKDP	PLFYLD	AQKGR	624		
mmFATP4_592	GT	F	KFQKTEL	RKEG	FDPSVVKDP	PLFYLD	ARKGC	624		
hsFATP1_595	GT	F	KTIQKT	RLQR	REGFDP	RQTS	DRFL	FLDLKQGH	627	
hsFATP4_625	YVPLDQ	EAYS	RIQAGE	EKKL					643	
mmFATP4_625	YV	ALDQ	EAYTRI	QAGE	EKKL				643	
hsFATP1_628	Y	LPLNE	AVYTRI	CS	G	A	F	A	L	646

FIG. 39C

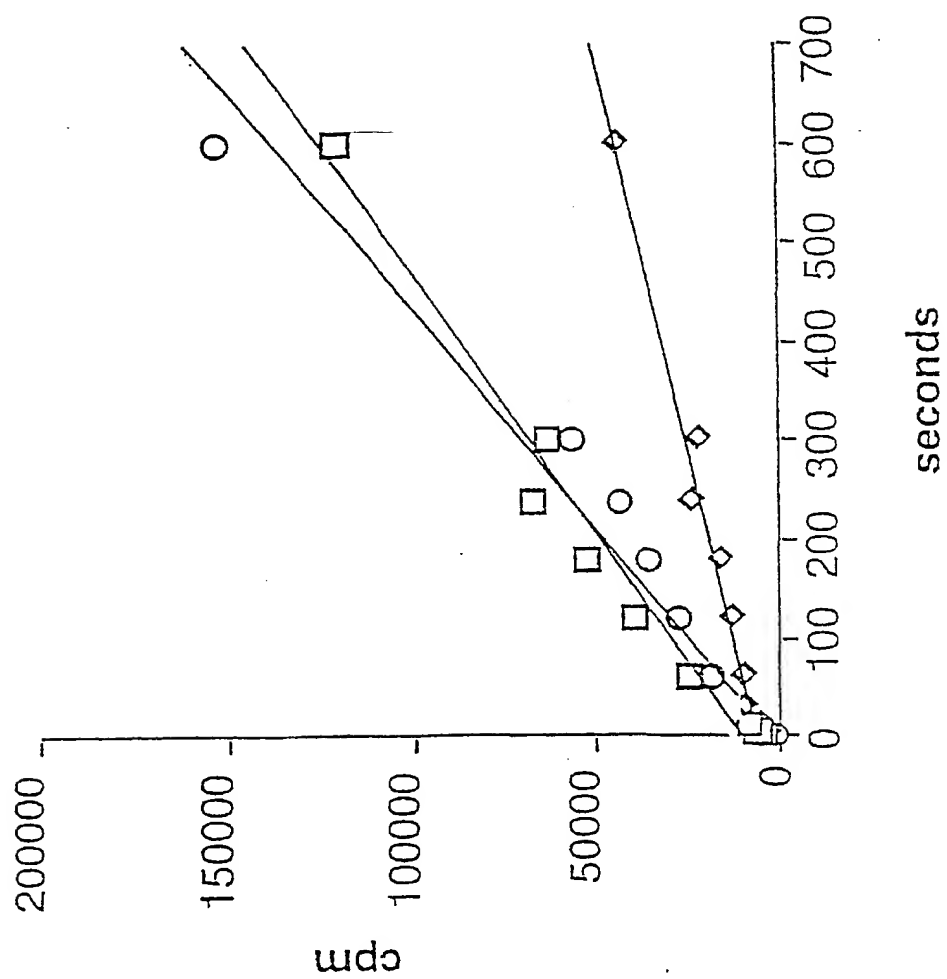


FIG. 40

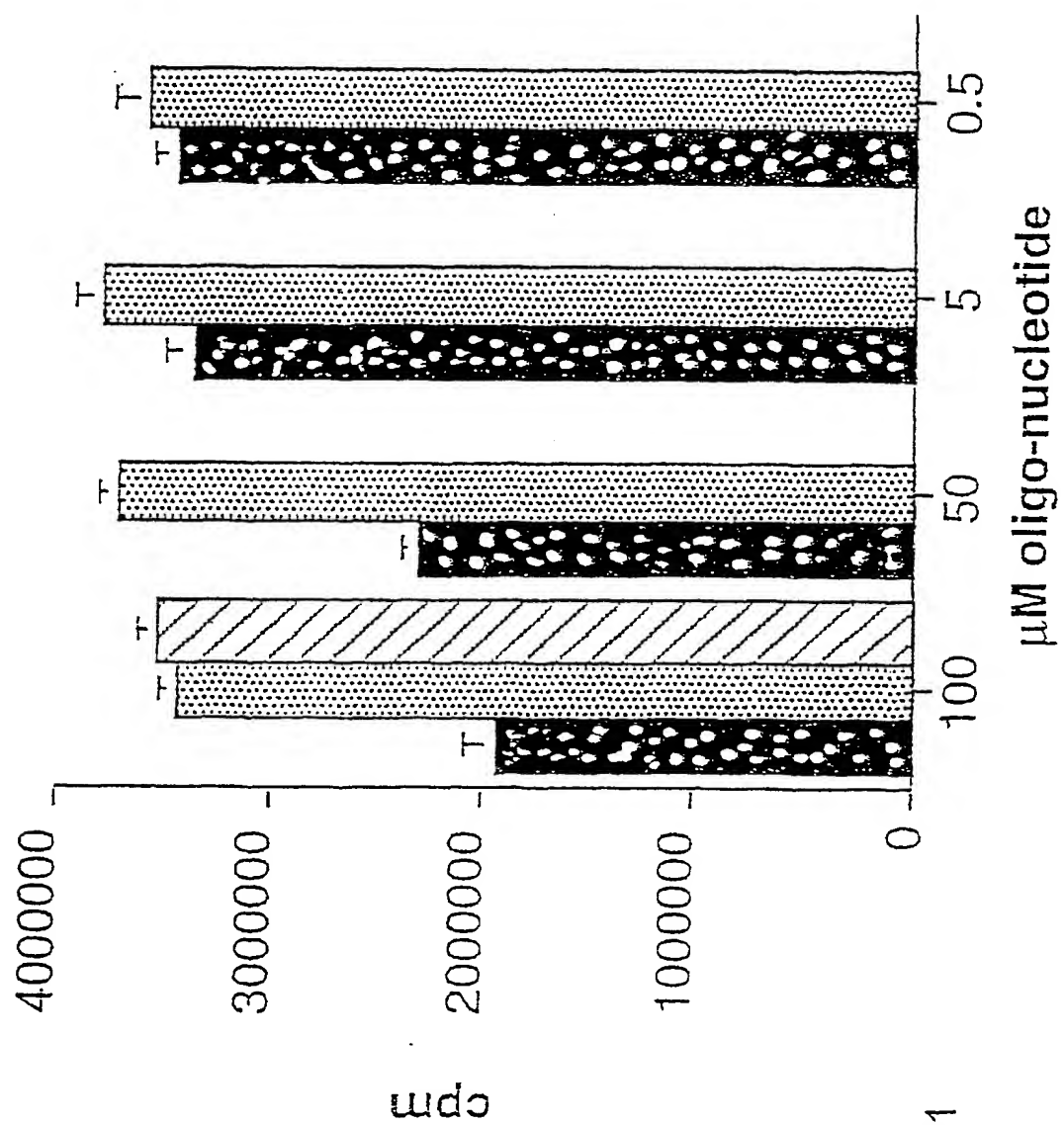


FIG. 41

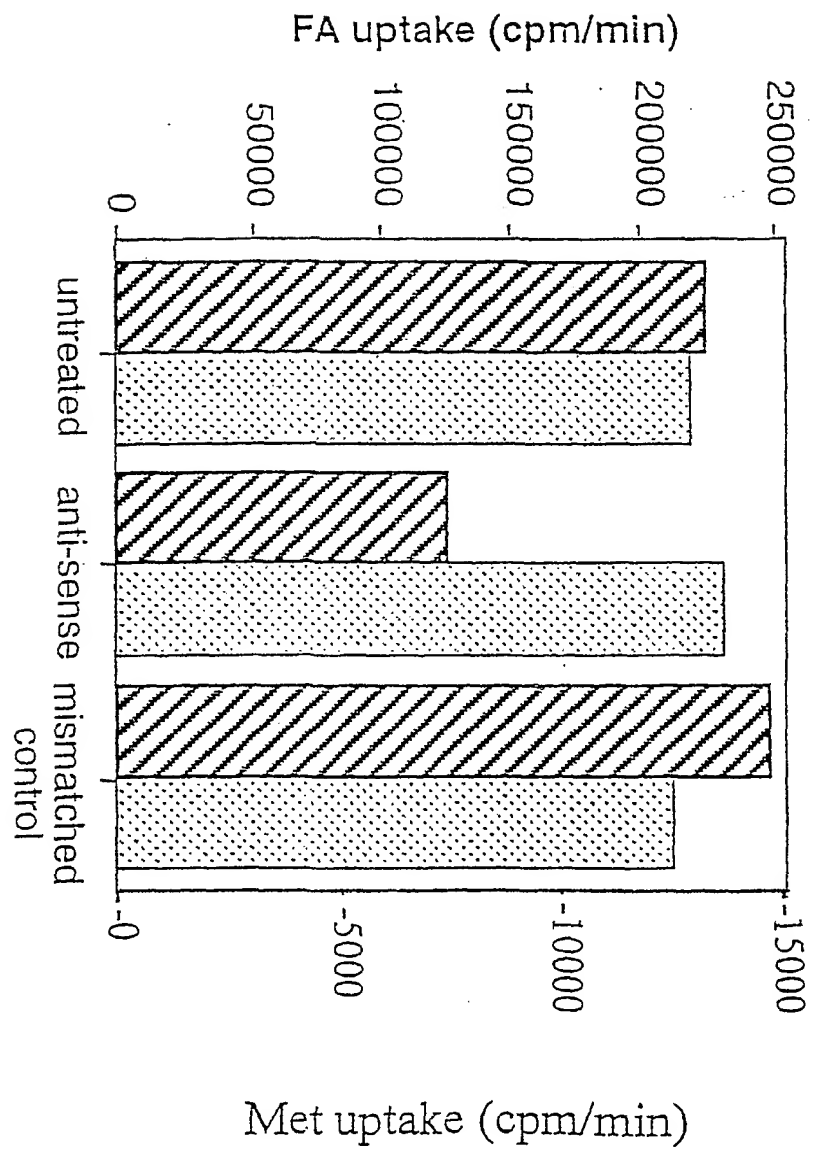


FIG. 42

ATGCTGCTTGGAGCCTCTCTGGTGGGGGCGCTACTGTTCTCCAAGC  
TAGTGCTGAAGCTGCCCTGGACCCAGGTGGGATTCTCCCTGTTGCT  
CCTGTACTTGGGGTCTGGTGGCTGGCGTTTCATCCGGGTCTTCATC  
AAGACGGTCAGGAGAGATATCTTTGGTGGCATGGTGCTCCTGAAGG  
TGAAGACCAAGGTGCGACGGTACCTTCAGGAGCGGAAGACGGTGCC  
CCTGCTGTTTGCTTCAATGGTACAGCGCCACCCGGACAAGACAGCC  
CTGATTTTTCGAGGGCACAGACACTCACTGGACCTTCCGCCAGCTGG  
ATGAGTACTCCAGTAGTGTGGCCAACTTCCTGCAGGCCCGGGGCCCT  
GGCCTCAGGCAATGTAGTTGCCCTCTTTATGGAAAACCGCAATGAG  
TTTGTGGGTCTGTGGCTAGGCATGGCCAAGCTGGGCGTGGAGGCGG  
CTCTCATCAACACCAACCTTAGGCGGGATGCCCTGCGCCACTGTCT  
TGACACCTCAAAGGCACGAGCTCTCATCTTTGGCAGTGAGATGGCC  
TCAGCTATCTGTGAGATCCATGCTAGCCTGGAGCCCACTCAGCC  
TCTTCTGCTCTGGATCCTGGGAGCCCAGCACAGTGCCCGTCAGCAC  
AGAGCATCTGGACCCTCTTCTGGAAGATGCCCCGAAGCACCTGCCC  
AGTCACCCAGACAAGGGTTTTACAGATAAGCTCTTCTACATCTACA  
CATCGGGCACCACGGGGCTACCCAAAGCTGCCATTGTGGTGCACAG  
CAGGTATTATCGTATGGCTTCCCTGGTGTACTATGGATTCCGCATG  
CGGCCTGATGACATTGTCTATGACTGCCTCCCCCTCTACCACTCAA  
GCAGGAAACATCGTGGGGATTGGCAGTGCTTACTCCACGGCATGAC  
TGTGGTGATCCGGAAGAAGTTCTCAGCCTCCCGGTCTGGGATGAT  
TGTATCAAGTACAACCTGCACAGTGGTACAGTACATTGGCGAGCTCT  
GCCGCTACCTCCTGAACCAGCCACCCCGTGAGGCTGAGTCTCGGCA  
CAAGGTGCGCATGGCACTGGGCAACGGTCTCCGGCAGTCCATCTGG  
ACCGACTTCTCCAGCCGTTTCCACATCCCCCAGGTGGCTGAGTTCT  
ATGGGGGCCACTGAATGCAACTGTAGCCTGGGCAACTTTGACAGCCG  
GGTGGGGGCCTGTGGCTTCAATAGCCGCATCCTGTCCTTTGTGTAC  
CCTATCCGTTTGGTACGTGTCAATGAGGATAACCATGGAACCTGATCC  
GGGGACCCGATGGAGTCTGCATTCCCTGTCAACCAGGTCAGCCAGG  
CCAGCTGGTGGGTCCCATCATCCAGCAGGACCCTCTGCGCCGTTTC

FIG. 43A

GACGGGTACCTCAACCAGGGTGCCAACAACAAGAAGATTGCTAATG  
ATGTCTTCAAGAAGGGGGACCAAGCCTACCTCACTGGTGACGTCCT  
GGTGATGGATGAGCTGGGTTACCTGTACTTCCGAGATCGCACTGGG  
GACACGTTCCGCTGCAAAGGGGAGAATGTATCTACCACTGAGGTGG  
AGGGCACACTCAGCCGCCTGCTTCATATGGCAGATGTGGCAGTTTA  
TGGTGTGAGGTGCCAGGAAGTGAAGGCCGAGCAGGAATGGCTGCC  
GTTGCAAGTCCCATCAGCAACTGTGACCTGGAGAGCTTTGCACAGA  
CCTTGAAAAAGGAGCTGCCTCTGTATGCCCCGCCCATCTTCCTGCG  
CTTCTTGCTGAGCTGCACAAGACAGGGACCTTCAAGTTCCAGAAG  
ACAGAGTTGCGGAAGGAGGGCTTTGACCCATCTGTTGTGAAAGACC  
CGCTGTTCTATCTGGATGCTCGGAAGGGCTGCTACGTTGCACTGGA  
CCAGGAGGCCTATACCCGCATCCAGGCAGGCGAGGAGAAGCTGTGA  
TTTCCCCCTACATCCCTCTGAGGGCCAGAAGATGCTGGATTGAGAG  
CCCTAGCGTCCACCCCAAGAGGGTCTTGGGCAATGCCAGACCAAAGC  
TAGCAGGGCCCGCACCTCCGCCCCCTAGGTGCTGATCTCCCCCTCTCC  
CAAAGTCCAAGTGACTCACTGCCGCTTCCCCGACCCTCCAGAGGC  
TTTCTGTGAAAGTCTCATCCAAGCTGTGTCTTCTGGTCCAGGCGTG  
GCCCCCTGGCCCCAGGGTTTCTGATAGGCTCCTTTAGGATGGTATCT  
TGGGTCCAGCGGGCCAGGGTGTGGGAGAGGAGTCACTAAGATCCCT  
CCAATCAGAAGGGAGCTTACAAAGGAACCAAGGCAAAGCCTGTAGA  
CTCAGGAAGCTAAGTGGCCAGAGACTATAGTGGCCAGTCATCCCAT  
GTCCACAGAGGATCTTGGTCCAGAGCTGCCAAAGTGTCACCTCTCC  
CTGCCTGCACCTCTGGGGAAAAGAGGACAGCATGTGGCCACTGGGC  
ACCTGTCTCAAGAAGTCAGGATCACACACTCAGTCCTTGTTTCTCC  
AGGTTCCTTGTTCTTGTCTCGGGGAGGGAGGGACGAGTGTCTGT  
CTGTCCTTCCTGCCTGTCTGTGAGTCTGTGTTGCTTCTCCATCTGT  
CCTAGCCTGAGTGTGGGTGGAACAGGCATGAGGAGAGTGTGGCTCA  
GGGGCCAATAAACTCTGCCTTGACTCCTCTTAAAAAAAAAAAAAAAA  
AA

FIG. 43B

MLLGASLVGALLFSKLVKL PWTQVGFSLLLLLYLGSGGWRFIRVFI  
KTVRRDIFGGMVLLKV KTKVRRYLQERKTVPLL FASMVQRHPDKTA  
LIFEGTDTHWTFRQLDEYSSSVANFLQARGLASGNVVALFMENRNE  
FVGLWLGMAKLGVEAALINTNLRRDALRHCLDTSKARALIFGSEMA  
SAICEIHASLEPTLSLFCSGSWEPTVPVSTEHLDPILLEDAPKHL P  
SHPDKGFTDKLFYIYTSGTTGLPKAAIVVHSRYYRMASLVYYGFRM  
RPDDIVYDCLPLYHSSRKHRGDWQCLLHGMTVVIRKKFSASRFWDD  
CIKYNCTVVQYIGELCRYLLNQPPREAESRHKVRMALGNGLRQSIW  
TDFSSRFHIPQVAEFYGATECNC SLGNFDSRVGACGFNSRILSFVY  
PIRLVRVNEDTMELIRGPDGVCIPCQPGQPGQLVGRI IQQDPLRRF  
DGYLNQGAN NKKIANDVFKKGDQAYLTGDVLVMDELGYLYFRDRTG  
DTFRWKGENVSTTEVEGTLSRLLHMADVAVYGVEVPGTEGRAGMAA  
VASPISNCDLESFAQTLKKELPLYARPIFLRFLPELHKTGT FKFQK  
TEL RKEGFDPSVVKDPLFYLDARKGCYVALDQEAYTRI QAGEEKL

FIG. 43C



10 20 30 40

TCGACCCACGGCGTCCGGGACCCCAAAGCAGAAGCCCGCA 40  
CAGTAGGCACAGCGCACCCAAGAAGGGTCCAGGAGTCTGC 80  
AGAAACAGAAAGGTCCCCGGCCTCAGCCTCCTAGTCCCTG 120  
CCTGCCTCCTGCCTGAGCTTCTGGGAGACTGAAGGCACGG 160  
CTTGCAGCTTCAGGATGCGGGCTCCGGGTGCGGGCGCGGC 200

210 220 230 240

CTCGGTGGTCTCGCTGGCGCTGTTGTGGCTGCTGGGGGCTG 240  
CCGTGGACCTGGAGCGCGGCAGCGGCGCTCGGCGTGTACG 280  
TGGGCAGCGGCGGCTGGCGCTTCTCGGCATCGTCTGCAA 320  
GACCGCGAGGCGAGACCTCTTCGGTCTCTCTGTGCTGATC 360  
CGCGTGCGCCTGGAGCTGCGGCGGCACCAGCGTGCCGGCC 400

410 420 430 440

ACACCATCCCGCGCATCTTTCAGGCGGTAGTGCAGCGACA 440  
GCCCGAGCGCCTGGCGCTGGTGGATGCCGGGACCGGCGAG 480  
TGCTGGACCTTTGCGCAGCTGGACGCCTACTCCAATGCGG 520  
TAGCCAACCTCTTCCGCCAGCTGGGCTTTCGCGCCGGGCGA 560  
CGTGGTGGCCATCTTCCTGGAGGGCCGGCCGGAGTTCGTG 600

610 620 630 640

GGGCTGTGGCTGGGCCTGGCCAAGGCGGGCATGGAGGCCG 640  
CGCTGCTCAACGTGAACCTGCGGCGCGAGCCCTTGGCCTT 680  
CTGCCTGGGCACCTCGGGCGCTAAGGCCCTGATCTTTGGA 720  
GGAGAAATGGTGGCGGCGGTGECGAAGTGAGCGGGCATC 760  
TGGGGAAAAGTTTGATCAAGTTCTGCTCTGGAGACTTGGG 800

810 820 830 840

GCCCGAGGGCATCTTGCCGGACACCCACCTCCTGGACCCG 840  
CTGCTGAAGGAGGCCTCTACTGCCCCCTTGGCACAGATCC 880  
CCAGCAAGGGCATGGACGATCGTCTTTTCTACATCTACAC 920  
GTGCGGGGACCACCGGGCTGCCCAAGGCTGCCATTGTCTGT 960  
CACAGCAGGTACTACCGCATGGCAGCCTTCGGCCACCACG 1000

1010 1020 1030 1040

CCTACCGCATGCAGGCGGCTGACGTGCTCTATGACTGCCT 1040  
GCCCTGTACCACTCGGCAGGAAACATCATCGGCGTGGGG 1080  
CAGTGTCTCATCTATGGGCTGACAGTCGTCTCCGCAAGA 1120  
AATTCTCGGCCAGCCGCTTCTGGGACGACTGCATCAAGTA 1160  
CAACTGCACGGTGGTTTCACTACATCGGGGAGATCTGCCGC 1200

FIG. 44A

1210	1220	1230	1240
TACCTGCTGAAGCAGCCGGTGCGCGAGGCGGAGAGGCGAC 1240			
ACCGCGTGCGCCTGGCGGTGGGGAACGGGCTGCGTCCTGC 1280			
CATCTGGGAGGAGTTTCACGGAGCGCTTCGGCGTACGCCAA 1320			
ATCGGGGAGTTCTACGGCGCCACCGAGTGCAACTGCAGCA 1360			
TTGCCAACATGGACGGCAAGGTCGGCTCCTGTGGTTTCAA 1400			
1410	1420	1430	1440
CAGCCGCATCCTGCCCCACGTGTACCCCATCCGGCTGGTG 1440			
AAGGTCAATGAGGACACAATGGAGCTGCTGCGGGATGCCC 1480			
AGGGCCTCTGCATCCCCTGCCAGGCCGGGGAGCCTGGCCT 1520			
CCTTGTGGGTCAGATCAACCAACAGGACCCGCTGCGCCGC 1560			
TTCGATGGCTATGTCAGCGAGAGCGCCACCAGCAAGAAGA 1600			
1610	1620	1630	1640
TCGCCCACAGCGTCTTCAGCAAGGGCGACAGCGCCTACCT 1640			
CTCAGGTGACGTGCTAGTGATGGATGAGCTGGGCTACATG 1680			
TACTTCCGGGACCGTAGCGGGGACACCTTCCGCTGGCGAG 1720			
GGGAGAACGTCTCCACCACCGAGGTGGAGGGCGTGCTGAG 1760			
CCGCCTGCTGGGCCAGACAGACGTGGCCGTCTATGGGGTG 1800			
1810	1820	1830	1840
GCTGTTCCAGGAGTGGAGGGTAAGGCAGGGATGGCGGCCG 1840			
TCGCAGACCCCCACAGCCTGCTGGACCCCCAACGCGATATA 1880			
CCAGGAGCTGCAGAAGGTGCTGGCACCCCTATGCCCGGCC 1920			
ATCTTCCTGCGCCTCCTGCCCCAGGTGGACACCACAGGCA 1960			
CCTTCAAGATCCAGAAGACGAGGCTGCAGCGAGAGGGCTT 2000			
2010	2020	2030	2040
TGACCCACGCCAGACCTCAGACCGGCTCTTCTTCCTGGAC 2040			
CTGAAGCAGGGCCACTACCTGCCCTTAAATGAGGCAGTCT 2080			
ACACTCGCATCTGCTCGGGCGCCTTCGCCCTCTGAAGCTG 2120			
TTCCTCTACTGGCCACAACTCTGGGCCTGGTGGGAGAGG 2160			
CCAGCTTGAGCCAGACAGCGCTGCCAGGGGTGGCCGCCT 2200			

FIG. 44B

2610 2620 2630 2640

GGTCAGGCTGGTCTTGAACCTCCTGACCTCAGGTGATCCGC 2640  
TGGCCTCGGCCTCCCAGAGTGCTGGGATTATAGGCGTGAG 2680  
CCTCTGGCCCCGGCCTTTTCTTTTTCTCTCTCTCTCTGCC 2720  
GAGAGTGGAACACACGTGTCTGGGAGCTGCATCTTGTGT 2760  
AGGGTCCAGCTGCTTTTGGGGACTGCAGGAATCATCTCCC 2800

2810 2820 2830 2840

CTGGGGCCCTGGACTCGGACTGGGGCCTCCCCACCTCCCTC 2840  
TCGGCTGTGCCTTACGGAGCCCCAATCCAGGCCTCCTGTG 2880  
GCTGTTGGGTTCAGATGCTGCAGCTCCATGTGACTTCCA 2920  
AGCAGGCCCTCCGCCCTCCCTGCTGAATGGAGGAGCCGGG 2960  
GGTCCCCCAGGCCAACTGGAAAATCTCCCAGGCTAGGCCA 3000

3010 3020 3030 3040

ATTGCCTTTTGCACCTTCCCCGTTCTGTACATTTCCCCA 3040  
GCCCCACCTTCCCCCTCCTGATGCCCTGAAAGCTTCCGGAA 3080  
TTGACTGTGACCACTTGGATGTCACCACTGTCAGCCCCCTG 3120  
CCTTGATGTCCCCATTTAGCCATCTCCATGGAGCTCCTGC 3160  
TGGAGGGCCCTGAACCTGCACTGCGTGGCTGCCAGCCA 3200

3210 3220 3230 3240

GCTGCCTCCTGTCCTGGGAGGAGGCCTCCTGGGTGTCCTC 3240  
ATCTGGTGTGTCTACTGGAGGGTCCCACAGGAGAGGCAGC 3280  
AGAGGGGTCAGGGGAGGTCTCCTGCCGGGGGTGGCCTCT 3320  
CAAGCCTCAGGGGTTCTAGCCTGTTGAATATACCCACCT 3360  
GGTGGGTGGCCCCCTCCGATGTCCCCACTGATGGCTCTGAC 3400

3410 3420 3430 3440

ACCGTGTTGGTGGCGATGTCCCAGACAATCCCACCAGGAC 3440  
GGCCCAGACATCCCTACTGGCTTCGCTGGTGGCTCATCTC 3480  
GAACATCCACGCCAGCCTTTCTGGGGCCGGCCACCCAGGC 3520  
CGCCTGTCCGTCTGTCTCCTCCAGCAGCACCCCTGGC 3560  
CCCTGGAGTGGTGGGGCCATGGCAAGAGACACCGTGGCGT 3600

3610 3620 3630 3640

CTCATGTGAACCTTCTCTGGGCACTGTGGTTTTATTTCTTA 3640  
ATTGATTTAAGAAATAAACCTGAAGACCGTCTGGTGAAAA 3680  
AAAAAAAAAAAAAAAAA 3694

FIG. 44C

2610	2620	2630	2640
..... ..... ..... .....			
GGTCAGGCTGGTCTTGAACCTCCTGACCTCAGGTGATCCGC 2640			
TGGCCTCGGCCTCCCAGAGTGCTGGGATTATAGGCCTGAG 2680			
CCTCTGGCCCCGGCCTTTTCCTTTTTTCCTCTCCTCTCCTGCC 2720			
GAGAGTGGAACACACGTGTCCTGGGAGCTGCATCTTGTGT 2760			
AGGGTCCAGCTGCTTTTGGGGACTGCAGGAATCATCTCCC 2800			
2810	2820	2830	2840
..... ..... ..... .....			
CTGGGCCCTGGACTCGGACTGGGGCCTCCCCACCTCCCTC 2840			
TCGGCTGTGCCTTACGGAGCCCCAATCCAGGCCTCCTGTG 2880			
GCTGTTGGGTTCCAGATGCTGCAGCTCCATGTGACTTCCA 2920			
AGCAGGCCCTCCGCCCTCCCTGCTGAATGGAGGAGCCGGG 2960			
GGTCCCCCAGGCCAACTGGAAAATCTCCCAGGCTAGGCCA 3000			
3010	3020	3030	3040
..... ..... ..... .....			
ATTGCCTTTTGCACCTTCCCCGTTTCCTGTCACATTTCCCCA 3040			
GCCCCACCTTCCCCCTCCTGATGCCCTGAAAGCTTCCGGAA 3080			
TTGACTGTGACCACTTGGATGTCACCACTGTCAGCCCCCTG 3120			
CCTTGATGTCCCCATTTAGCCATCTCCATGGAGCTCCTGC 3160			
TGGAGGGCCCTGAACCCCTGCACTGCGTGECTGCCCAGCCA 3200			
3210	3220	3230	3240
..... ..... ..... .....			
GCTGCCTCCTGTCTCCTGGGAGGAGGCCTCCTGGGTGTCTC 3240			
ATCTGGTGTGTCTACTGGAGGGTCCCACAGGAGAGGCAGC 3280			
AGAGGGGTCAGGGGAGGTCTCCTGCCGGGGGTTGGCCTCT 3320			
CAAGCCTCAGGGGTTCAGCCTGTTGAATATACCCACCT 3360			
GGTGGGTGGCCCCCTCCGATGTCCCCACTGATGGCTCTGAC 3400			
3410	3420	3430	3440
..... ..... ..... .....			
ACCGTGTTGGTGGCGATGTCCCAGACAATCCCACCAGGAC 3440			
GGCCCAGACATCCCTACTGGCTTTCCTGGTGGCTCATCTC 3480			
GAACATCCACGCCAGCCTTTCTGGGGCCGGCCACCCAGGC 3520			
CGCCTGTCCGTCTGTCTCCTCCCTCCAGCAGCACCCCTGGC 3560			
CCCTGGAGTGGTGGGGCCATGGCAAGAGACACCGTGGCGT 3600			
3610	3620	3630	3640
..... ..... ..... .....			
CTCATGTGAACTTTCTTGGGCACTGTGGTTTTATTTCCTA 3640			
ATTGATTTAAGAAATAAACCTGAAGACCGTCTGGTGAAAA 3680			
AAAAAAAAAAAAAAAA 3694			

FIG. 44D

10 20 30 40  
 MRAPGAGAASVVSLLALLWLLGLPWTWSAAAALGVYVGS GG 40  
 WRFLRIVCKTARRDLFGLSVLIRVRLELRRHRAGHTIPR 80  
 IFQAVVQRQPERLALVDAGTGECWTFQAQLDAYSNVANLF 120  
 RQLGFAPGDVVAIFLEGRPEFVGLWLGLAKAGMEAALLNV 160  
 NLRREPLAFCLGTSGAKALIFGGEMVAAVA EVSGHLGKSL 200  
 210 220 230 240  
 IKFCSGDLGPEGILPDTHLLDPLLKEASTAPLAQIPSKGM 240  
 DDRLFYIYTS GTTGLPKAAIVVHSRYRMAAFGHHAYR MQ 280  
 AADVLYDCLPLYHSAGNIIGVGQCL IYGLTVVLRKKFSAS 320  
 RFWDDCIKYNCTVVQYIGEICRYLLKQPVREAERRHRVRL 360  
 AVGNGLRPAIWE EFTERFGVRQIGEFYGATECNC SIANMD 400  
 410 420 430 440  
 GKVGSCGFNSRILPHVYPIRLVKVNEDTMELLRDAQGLCI 440  
 PCQAGEPGLLVGQINQQDPLRRFDGYVSESATSKKIAHSV 480  
 FSKGDSAYLSGDVLVMDELGYMYFRDRSGDTFRWRGENVS 520  
 TTEVEGVLSRLLGQTDVAVYGVAVPGVEGKAGMAAVADPH 560  
 SLLDPNAIYQELQKVLAPYARPIFLRLLPQVDTTGTFKIQ 600  
 610 620 630 640  
 KTRLQREGFDPRQTS DRLFFLDLKQGHYLP LNEAVYTRIC 640  
 SGAFAL 646

FIG. 45

10	20	30	40
-----			
GGAATTCCAAAAAAAAAAAAATACGACTACACCTGCTCCGG	40		
AGCCCGCGGGCGGTACCTGCAGCGGAGGAGCTCTGTCTTCC	80		
CCTTCATCTCACGCGAGCCCGGCGTCCCGCCGCGTGCGCC	120		
CCGGCGCAGCCCGCCAGTCCGCCCCGAGCCCGCCAGTCG	160		
CCGCGCTGCACGCCCCGGGGTGAACCCTCTGCCCTCGCTGG	200		
210	220	230	240
-----			
GACAGAGGGCCCCGCGAGCCGTCATGCTTTCCGCCATCTAC	240		
ACAGTCCTGGCGGGACTGCTGTTCTCGCCGCTCCTGGTGA	280		
ACCTCTGCTGCCCATACTTCTTCCAGGACATAGGCTACTT	320		
CTTGAAGGTGGCCGCGGTGGGCCGGAGGGTGCGCAGCTAC	360		
GGGCAGCGGCGGCCGCGCGCACCATCCTGCGGGCGTTCC	400		
410	420	430	440
-----			
TGGAGAAAGCGCGCCAGACGCCACACAAGCCTTTTCTGCT	440		
CTTCCGCGACGAGACTCTCACCTACGCGCAGGTGGACCGG	480		
CGCAGCAATCAAGTGGCCCCGGGCGCTGCACGACCACCTCG	520		
GCCTGCGCCAGGGAGACTGCGTGCGCTCCTTATGGGTAA	560		
CGAGCCGGCCTACGTGTGGCTGTGGCTGGGGCTGGTGAAG	600		
610	620	630	640
-----			
CTGGGCTGTGCCATGGCGTGCCTCAATTACAACATCCGCG	640		
CGAAGTCCCTGCTGCACTGCTTCCAGTGCTGCGGGGCGAA	680		
GGTGCTGCTGGTGTGCGCCAGAACTACAAGCAGCTGTGAA	720		
GAGATACTGCCAAGCCTTAAAAAAGATGATGTGTCCATCT	760		
ATTATGTGAGCAGAACTTCTAACACAGATGGGATTGACTC	800		
810	820	830	840
-----			
TTTCCTGGACAAAGTGGATGAAGTATCAACTGAACCTATC	840		
CCAGAGTCATGGAGGTCTGAAGTCACTTTTTTCCACTCCTG	880		
CCTTATACATTTATACTTCTGGAACCACAGGTCTTCCAAA	920		
AGCAGCCATGATCACTCATCAGCGCATATGGTATGGAACT	960		
GGCCTCACTTTTGTAAAGCGGATTGAAGGCAGATGATGTCA	1000		
1010	1020	1030	1040
-----			
TCTATATCACTCTGCCCTTTTACCACAGTGCTGCACTACT	1040		
GATTGGCATTACGGATGTATTGTGGCTGGTGCTACTCTT	1080		
GCCTTGCGGACTAAATTTTCAGCCAGCCAGTTTTGGGATG	1120		
ACTGCAGAAAATACAACGTCACTGTCAATTCAGTATATCGG	1160		
TGAACTGCTTCGGTATTTATGCAACTCACCACAGAAACCA	1200		

FIG. 46A

1210	1220	1230	1240
AATGACCGTGATCATAAAGTGAGACTGGCACTGGGAAATG 1240			
GCTTACGAGGAGATGTGTGGAGACAATTTGTCAAGAGATT 1280			
TGGGGACATATGCATCTATGAGTTCTATGCTGCCACTGAA 1320			
GGCAATATTGGATTTATGAATTATGCGAGAAAAGTTGGTG 1360			
CTGTTGGAAGAGTAAACTACCTACAGAAAAAAATCATAAC 1400			
1410	1420	1430	1440
TTATGACCTGATTAAATATGATGTGGAGAAAGATGAACCT 1440			
GTCCGAGATGAAAATGGATATTGCGTCAGAGTTCCCAAAG 1480			
GTGAAGTTGGACTTCTGGTTTGCAAAATCACACAACTTAC 1520			
ACCATTTAATGGCTATGCTGGAGCAAAGGCTCAGACAGAG 1560			
AAGAAAAAACTGAGAGATGTCTTTAAGAAAGGAGACCTCT 1600			
1610	1620	1630	1640
ATTTCAACAGTGGAGATCTCTTAATGGTTGACCATGAAAA 1640			
TTTCATCTATTTCCACGACAGAGTTGGAGATACATTCCGG 1680			
TGGAAGGGGGAAAATGTGGCCACCACTGAAGTTGCTGATA 1720			
CAGTTGGACTGGTTGATTTTGTCCAAGAAGTAAATGTTTA 1760			
TGGAGTGCATGTGCCAGATCATGAGGGTCCGATTGGCATG 1800			
1810	1820	1830	1840
GCCTCCATCAAAATGAAAGAAAACCATGAATTTGATGGAA 1840			
AGAAACTCTTTTCAGCACATTGCTGATTACCTACCTAGTTA 1880			
TGCAAGGCCCCGGTTTCTAAGAATACAGGACACCATTGAG 1920			
ATCACTGGAACCTTTTAAACACCGCAAAATGACCCTGGTGG 1960			
AGGAGGGCTTTAACCCTGCTGTCATCAAAGATGCCTTGTA 2000			
2010	2020	2030	2040
TTTCTTGGATGACACAGCAAAAATGTATGTGCCTATGACT 2040			
GAGGACATCTATAATGCCATAAGTGCTAAAACCCTGAAAC 2080			
TCTGAATATTCCCAGGAGGATAACTCAACATTTCCAGAAA 2120			
GAAACTGAATGGACAGCCACTTGATATAATCCAACCTTTAA 2160			
TTTGATTGAAGATTGTGAGGAAATTTTGTAGGAAATTTGC 2200			
2210	2220	2230	2240
ATACCCGTAAAGGGGAGACTTTTTTTAAATAACAGTTGAGTC 2240			
TTTGCAAGTAAAAAGATTTAGAGATTATTATTTTTTCAGTG 2280			
TGCACCTACTGTTTGTATTTGCAAACTGAGCTTGTTGGAG 2320			
GGAAGGCATTATTTTTTTAAATACTTAGTAAATTAAATGA 2360			
AC 2362			

FIG. 46B

**SUBSTITUTE SHEET (RULE 26)**



10	20	30	40
AAGTTCTCGGCTGGTCAGTTCTGGGAAGATTGCCAGCAGC 40			
ACAGGGTGACGGTGTTCCAGTACATTGGGGAGCTGTGCCG 80			
ATACCTTGTCAACCAGCCCCGAGCAAGGCAGAACGTGGC 120			
CATAAGGTCCGGCTGGCAGTGGGCAGCGGGCTGCGCCAG 160			
ATACCTGGGAGCGTTTTGTGCGGCGCTTCGGGCCCTGCA 200			
210	220	230	240
GGTGCTGGAGACATATGGACTGACAGAGGGCAACGTGGCC 240			
ACCATCAACTACACAGGACAGCGGGGCGCTGTGGGGCGTG 280			
CTTCCTGGCTTTTACAAGCATATCTTCCCTTCTCCTTGAT 320			
TCGCTATGATGTCACCACAGGAGAGCCAATTTCGGGACCCC 360			
CAGGGGCACTGTATGGCCACATCTCCAGGTGAGCCAGGGC 400			
410	420	430	440
TGCTGGTGGCCCCGGTAAGCCAGCAGTCCCCATTCTCTGGG 440			
CTATGCTGGCGGGGCCAGAGCTGGCCCAGGGGAAGTTGCTA 480			
AAGGATGTCTTCCGGCCTGGGGATGTTTTCTTCAACACTG 520			
GGGACCTGCTGGTCTGCGATGACCAAGGTTTTCTCCGCTT 560			
CCATGATCGTACTGGAGACACCTTCAGGTGGAAGGGGGAG 600			
610	620	630	640
AATGTGGCCACAACCGAGGTGGCAGAGGTCTTCGAGGCCC 640			
TAGATTTTCTTCAGGAGGTGAACGTCTATGGAGTCACTGT 680			
GCCAGGGCATGAAGGCAGGGCTGGAATGGCAGCCCTAGTT 720			
CTGCGTCCCCCCCACGCTTTGGACCTTATGCAGCTCTACA 760			
CCCACGTGTCTGAGAACTTGCCACCTTATGCCCGGCCCCG 800			
810	820	830	840
ATTCCTCAGGCTCCAGGAGTCTTTGGCCACCACAGAGACC 840			
TTCAAACAGCAGAAAGTTCCGGATGGCAAATGAGGGCTTCG 880			
ACCCCAGCACCCCTGTCTGACCCACTGTACGTTCTGGACCA 920			
GGCTGTAGGTGCCTACCTGCCCCCTCACAACCTGCCCGGTAC 960			
AGCGCCCTCCTGGCAGGAAACCTTCGAATCTGAGAACTTC 1000			
1010	1020	1030	1040
CACACCTGAGGCACCTGAGAGAGGAACTCTGTGGGGTGGG 1040			
GGCCGTTGCAGGTGTACTGGGCTGTCAGGGATCTTTTCTA 1080			
TACCAGAACTGCGGTCACATTTTTGTAATAAATGTGGCTG 1120			
GAGCTGATCCAGCTGTCTCTGACAAAAAAAAAAAAAAAAAA 1160			
AAAGGGCGGCCGC 1173			

FIG. 48

10	20	30	40
KFSAGQFWEDCQQHRVTVFQYIGELCRYLVNQPPSKAERG	40		
HKVRLAVGSGLRPDTWERFVRRFGPLQVLETYGLTEGNVA	80		
TINYTGORGAVGRASWLYKHIFPFSLIRYDVTGEPIRD	120		
QGHCMATSPGEPGLLVAPVSQOSPFLGYAGGPPELAQGKLL	160		
KDVFRPGDVFFNTGDLLVCDDQGFLRFHDRTGDTFRWKGE	200		
210	220	230	240
NVATTEVAEVFEALDFLQEVNVYGVTVPGHEGRAGMAALV	240		
LRPPHALDLMQLYTHVSENLPYARPRFLRLQESLATTET	280		
FKQOKVRMANEGFDPSTLSDPLYVLDQAVGAYLPLTTARY	320		
SALLAGNLR I	330		

FIG. 49

10 20 30 40  
 CGACCCACGCGTCCGGGCGGGCGGGGCCGGGCGGCGGGCG 40  
 GGGCTGGCGGGGCGGCCGGGCCATGCAGGGCGCAGAGCCG 80  
 GCTAAACCCTGCTGAGACCCGGCTCCGTGCGTCCAGGGGC 120  
 GGCTAATGCCCTCACGCTGTCTACGCTGCTGCAACCGGG 160  
 CCGCATCTGGACGGGGCGCCGCGGCGGAGCCGACGCCG 200  
 210 220 230 240  
 GGCCACAATGCTGCTTGGAGCCTCTCTGGTGGGGGTGCTG 240  
 CTGTTCTCCAAGCTGGTGCTGAAACTGCCCTGGACCCAGG 280  
 TGGGATTCTCCCTGTTGTTCTCTACTTGGGATCTGGCGG 320  
 CTGGCGCTTCATCCGGGTCTTCATCAAGACCATCAGGCGC 360  
 GATATCTTTGGCGGCCTGGTCCTCCTGAAGGTGAAGGCAA 400  
 410 420 430 440  
 AGGTGCGACAGTGCCCTGCAGGAGCGGCGGACAGTGCCCAT 440  
 TTTGTTTGCCTCTACCGTTTCGGCGCCACCCCGACAAGACG 480  
 GCCCTGATCTTCGAGGGCACAGATACCCACTGGACCTTCC 520  
 GCCAGCTGGATGAGTACTCAAGCAGTGTAGCCAACCTTCT 560  
 GCAGGCCCGGGGCGCTGGCCTCGGGCGATGTGGCTGCCATC 600  
 610 620 630 640  
 TTCATGGAGAACCQCAATGAGTTCGTGGGCCTATGGCTGG 640  
 GCATGGCCAAGCTCGGTGTGGAGGCAGCCCTCATCAACAC 680  
 CAACCTGCGGCGGGATGCTCTGCTCCACTGCCTCACCACC 720  
 TCGCGCGCACGGGCCCTTGTCTTTGGCAGCGAAATGGCCT 760  
 CAGCCATCTGTGAGGTCCATGCCAGCCTGGACCCCTCGCT 800  
 810 820 830 840  
 CAGCCTCTTCTGCTCTGGCTCCTGGGAGCCCGGTGCGGTG 840  
 CCTCCAAGCACAGAACACCTGGACCCTCTGCTGAAAGATG 880  
 CTCCCAAGCACCTTCCCAGTTGCCCTGACAAGGGCTTCAC 920  
 AGATAAACTGTTCTACATCTACACATCCGGCACCACAGGG 960  
 CTGCCCCAAGGCCGCCATCGTGGTGACACAGCAGGTATTACC 1000  
 1010 1020 1030 1040  
 GCATGGCTGCCCTGGTGTACTATGGATTCCGCATGCGGCC 1040  
 CAACGACATCGTCTATGACTGCCTCCCCCTCTACCACTCA 1080  
 GCAGGAAACATCGTGGGAATCGGCCAGTGCCCTGCTGCATG 1120  
 GCATGACGGTGGTGATTTCGGAAGAAGTTCTCAGCCTCCCG 1160  
 GTTCTGGGACGATTGTATCAAGTACAAGTGCACGATTGTG 1200

FIG. 50A

**SUBSTITUTE SHEET (RULE 26)**

2210	2220	2230	2240
AGCGGTCCTGGACAAGGCCAGACCAAAGCAAGCAGGGCCT	2240		
GGCACCTCCATCCTGAGGTGCTGCCCCTCCATCCAAAAC	2280		
GCCAAGTGACTCATTGCCTTCCCAACCCTTCCAGAGGCTT	2320		
TCTGTGAAAGTCTCATGTCCAAGTTCCGTCTTCTGGGCTG	2360		
GGCAGGCCCTCTGGTTCCCAGGCTGAGACTGACGGGTTTT	2400		

2410	2420	2430	2440
CTCAGGATGATGTCTTGGGTGAGGGTAGGGAGAGGACAAG	2440		
GGGTCACCGAGCCCTTCCCAGAGAGCAGGGAGCTTATAAA	2480		
TGGAACCAGAGCAGAAGTCCCCAGACTCAGGAAGTCAACA	2520		
GAGTGGGCAGGGACAGTGGTAGCATCCATCTGGTGGCCAA	2560		
AGAGAATCGTAGCCCCAGAGCTGCCCAAGTTCACTGGGCT	2600		

FIG. 50C

1

10 20 30 40

MLLGASLVGVLLFSKLVLKLPWTQVGFSLLFLYLGS GGWR 40  
FIRVFIKTIRRDIFGGLVLLKVAKVRQCLOERRTPILF 80  
ASTVRRHPDKTALIFEGTOTHWTFRQLDEYSSSVANFLQA 120  
RGLASGDVAAIFMENRNEFVGLWLGMAKLGVEAALINTNL 160  
RRDALLHCLTTSRARALVFGSEMASAICEVHASLDPSLSL 200

210 220 230 240

FCSGSWEPGA VPPSTEHLDP LLDAPKHL PSCPDKGFTDK 240  
LFYIYTS GTTGLPKAAIVVHSRYRMAALVYYGFRMRPND 280  
IVYDCLPLYHSAGNIVGIGQCL LHGMTVVIRKKFSASRFW 320  
DDCIKYNCTIVQYIGELCRYLLNQPPREAE NQHQVRMALG 360  
NGLRQSIWTFSSRFHIPQVAEFYGATECNC SLGNFDSQV 400

410 420 430 440

GACGFNSRILSFVYPIRLVRVNEDTMELIRGPDGVCIPCQ 440  
PGE PGQLVGRIIQKDPLRRFDGYLNQGANNKKIAKDVFKK 480  
GQOAYLTGDVLMDELGYLYFRDRTGDTFRWKGENVSTTE 520  
VEGTL SRLLDMADVAVYGVEVPGTEGRAGMAAVASPTGNC 560  
DLERFAQVLEKELPLYARPIFLRLLPELHKTGT YKFQKTE 600

610 620 630 640

LRKEGFDP AIVKDPLFYLDACKGRYVPLDQEAYSRIQAGE 640  
EKL 643

FIG. 51

GTCGTTGGGATCCTCGGCTGCTTAGATCTCGGAGCCACCTGTGTTCT  
GGCCCCCAAGTTCTCTACTTCCTGCTTCTGGGATGACTGTCGGCAGC  
ATGGCGTGACAGTGATCCTGTATGTGGGCGAGCTCCTGCGATACTTG  
TGTAACATTCCCCAGCAACCAGAGGACCGGACACATACAGTCCGCC  
TGGCAATGGGCAATGGACTACGGGCTGATGTGTGGGGGAGACCTTCC  
AGCAGCGTTTCGGTCCTATTTTCGGATCTNNGGAAGTCTTACGGGCTT  
CCACAGAAGGGCAACATGGGGCTTTAGTTCAAATATTGTTGGGGGC  
GCTGCGGGGGCCCTGGGGGCAAAGATGGAGCTTGCCTCCTCCGAATG  
CTGTCCCCCTTTGAGCTGGTGACAGTTCGACATGGAGGCGGCGGAGC  
CTGTGAGGGACAATCAGGGCTTCTGCATCCCTGTAGGGCTAGGGGA  
GCCGGGGCTGCTGTTGACCAAGGTGGTAAGCCAGCAACCCTTCGTG  
GGCTACCGCGGCCCCCGAGAGCTGTGCGGAACGGAAGCTGGTGCGCA  
ACGTGCGGCAATCGGGCGACGTTTACTACAACACCGGGGACGTACT  
GGCCATGGACCGCGAAGGCTTCCTCTACTTCCGCGACCGACTCGGG  
GACACCTTCCGATGGAAGGGCGAGAACGTGTCCACGCACGAGGTGG  
AGGGCGTGTTGTGCGCAGGTGGACTTCTTGCAACAGGTAAACGTGTAT  
GGCGTGTTGCGTGCCAGGTTGTGAGGGTAAGGTGGGCATGGCTGCTG  
TGGCATTAGCCCCCGGCCAGACTTTCGACGGGGAGAAGTTGTACCA  
GCACGTTGCGGCTTGGCTCCCTGCCTACGCTACCCCCCATTTTCATCC  
GCATCCAGGACGCCATGGAGGTCAACAGCACGTTCAAACCTGATGAA  
GACCCGGTTGGTGCGTGAGGGCTTCAATGTGGGGATCGTGGTTGAC  
CCTCTGTTTGTACTGGACAACCGGGCCCAGTCCTTCCGGCCCCCTGAC  
GGCAGAAATGTACCAGGCTGTGTGTGAGGGAACCTGGAGGCTCTGA  
TCACCTGGCCAACCCACTGGGGTAGGGATCAAAGCCAGCCACCCCC  
ACCCAACACACTCGGTGTCCCTTTTCATCCTGGGCCTGTGTGAATCC  
CAGCCTGGCCATACCCTCAACCTCAGTGGGCTGGAAATGACAGTGG  
GCCCTGTAGCAGTGGCAGAATAAACTCAGMTGYGTTACAGAAA

FIG. 52



	10	20	30	40	
VVGILGCLDLGATCVLAPKFSTSCFWDDCRQHGVTVILYV	40				
GELRLYLCNIPQQPEDRTHTVRLAMGNGLRADVWGDLPA	80				
FRSYFGSXEVLRASTEGOHGALVQILLGALRGPGGKDGAC	120				
LLRMLSPFELVQFDMEAAEPVRDNQGFVIPVGLGEPGLLL	160				
TKVVSQQPFVGYRGPRELSERKLVRNVRSQGDVYYNTGDV	200				
	210	220	230	240	
LAMDREGFLYFRDRLGDTFRWKGENVSTHEVEGVLSQVDF	240				
LQQVNVYGVCPGCEGKVGMAAVALAPGQTFDGEKLYQHV	280				
RAWLPAYATPHFIRIQDAMEVTSTFKLMKTRLVREGFNVG	320				
IVVDPLFVLDNRAQSFRPLTAEMYQAVCEGTWRL	354				

FIG. 53

**SUBSTITUTE SHEET (RULE 26)**

**Figure 1**

**PUBLISHED BY**

2210                      2220                      2230                      2240  
 AAGGAAGAGCAGGAATGGCTTCTATTATTTTAAAACCAA 2240  
 TACATCTTTAGATTTGGAAAAAGTTTATGAACAAGTTGTA 2280  
 ACATTTCTACCAGCTTATGCTTGTCCACGATTTTAAAGAA 2320  
 TTCAGGAAAAAATGGAAGCAACAGGAACATTCAAACATT 2360  
 GAAGCATCAGTTGGTGGGAAGATGGATTTAATCCACTGAAA 2400  
 2410                      2420                      2430                      2440  
 ATTTCTGAACCACTTTACTTTCATGGATAACTTGAAAAAGT 2440  
 CTTATGTTCTACTGACCAGGGAACTTTATGATCAAATAAT 2480  
 GTTAGGGGAAATAAACTTTAAGATTTTATATCTAGAAC 2520  
 TTTCATATGCTTTCTTAGGAAGAGTGAGAGGGGGGTATAT 2560  
 GATTCTTTATGAAATGGGGAAAGGGAGCTAACATTAATTA 2600

FIG. 54C

2610                      2620                      2630                      2640  
 TGCATGTACTATATTTCCCTTAATATGAGAGATAATTTTTT 2640  
 AATTGCATAAGAATTTTAATTTCTTTTAATTGATATAAAC 2680  
 ATTAGTTGATTATTCTTTTTATCTATTTGGAGATTCAGTG 2720  
 CATAACTAAGTATTTTCCTTAATACTAAAGATTTTAAATA 2760  
 ATAAATAGTGGCTAGCGGTTTGGACAATCACTAAAAATGT 2800  
 2810                      2820                      2830                      2840  
 ACTTTCTAATAAGTAAAATTTCTAATTTTGAATAAAAGAT 2840  
 TAAATTTTACTGAAAAAAAAAAAAAAAAAAAAAAAAATTGGCG 2880  
 GCCGC 2885

FIG. 54D

10	20	30	40
.....			
MLLSWLT	VLGAGMV	VLFLOKLL	FPYFWDDFWVLKVVLI 40
IIRLK	KEYEKR	GELVT	VLDKFLSHAKRQPRKPFIIYEGDIY 80
TYQDV	DKRSSR	VAHVFL	NHSSLKKGOTVALLMSNEPQFVH 120
VWFGL	AKLGCV	VAF	LNTNIRSN
LLNC	IRACG	PRAL	VVGA 160
DL	GTVEE	ILPS	LSENISVWGMKDSVPQGVISLKEKLSTS 200
210	220	230	240
.....			
PDEPV	PRSHH	VVSLL	KSTCLYIFTSGTTGLPKAAVISQLQ 240
VL	RGS	AVLW	AFGCTAHDIVYITLPLYHSSAAILGISGCVE 280
LG	ATC	VL	KKKFSASQFWSDCKKYDVTVFQYIGELCRYLCK 320
QSK	REG	EDHK	VRLAIGNGIRSDVWREFLDRFGNIKVCCEL 360
YA	ATESS	ISFM	NYTGRIGAIGRTNLFYKLLSTFDLIKDYDF 400
410	420	430	440
.....			
QKDEP	MRNE	QGWCI	HVKKGEPGLLISRVNAKNPFFFGYAGP 440
YKHTK	DKLL	CDV	FKKGDVYLNTGDLIVQDQDNFLYFWCRT 480
GD	FRW	KGEN	VATTEVADVIGMLDFIQEANVYGVAISGYE 520
GR	AGMAS	IILK	PNTSLDLEKVYEQVVTFLPAYACPRFLRI 560
QEK	MEAT	GT	FKLLKHOLVEDGFNPLKISEPLYFMDNLKKS 600
610	620	630	640
.....			
YVLLT	RELYDQ	IMLGE	IKL 619

FIG. 55

10	20	30	40
AAGTTCCCACTCCAGACTTCTGCGAGAACCCGTGAGGAAG 40			
CAGCGAGAACCGGGGGTTTGCAAGCCAGAGAAGGATGCGG 80			
ACTCCGGGAGCAGGAACAGCCTCTGTGGCCTCATTGGGGC 120			
TGCTTTGGCTTCTGGGACTTCCGTGGACCTGGAGCGCGGC 160			
GGCGGCGTTTCGGTGTGTACGTGGGTAGCGGTGGCTGGCGA 200			
210	220	230	240
TTTCTGCGTATCGTCTGCAAGACGGCGAGGCGAGACCTCT 240			
TTGGCCTCTCTGTCTGATCCGCGTGCGGCTAGAGCTACG 280			
ACGACACCGGGCGAGCAGGAGACACGATCCCACGCATCTTC 320			
CAGGCCGTGGCCCAGCGACACCGGAGCGCCTGGCGCTGG 360			
TAGATGCGAGTAGCGGTATCTGCTGGACCTTCGCACAGCT 400			
410	420	430	440
AGACACCTACTCCAATGCTGTGGCCAATCTGTTCCCTCCAG 440			
CTGGGCTTTGCGCCAGGCGATGTGGTGGCTGTGTTCCCTGG 480			
AAGGCCGGCCCGAGTTCGTGGGACTGTGGCTGGGCTTGGC 520			
CAAGGCCGGTGTAGTGGCTGCGCTTCTCAATGTCAACCTG 560			
AGGCCGGGAGCCCCCTTGCCCTTCTGCTTGGGCACATCAGCTG 600			
610	620	630	640
CCAAGGCCCTCATTATATGGCGGGGAGATGGCAGCGGCGGT 640			
GGCGGAGGTGAGTGAGCAGCTGGGGAAGAGCCTGCTCAAG 680			
TTCTGCTCTGGAGATCTGGGGCCTGAGAGCGTCCTGCCTG 720			
ACACGCAGCTTCTGGACCCCATGCTTGCTGAGGCGCCAC 760			
CACACCCCTGGCACAGGCCCCAGGCAAGGGCATGGATGAT 800			
810	820	830	840
CGGCTATTTTACATCTATACTTCTGGGACCACCGGACTTC 840			
CTAAGGCGGCCATTGTGGTGCACAGCAGGTACTACCGCAT 880			
CGCAGCCTTCGGCCACCATTCCTACAGCATGCGGGCCAAC 920			
GATGTGCTCTATGACTGCCTACCTCTCTACCACTCAGCAG 960			
GGAACATCATGGGCGTGGGACAGTGTATCATCTACGGGTT 1000			
1010	1020	1030	1040
AACGGTGGTACTGCGCAAGAAGTTCTCCGCCAGCCGCTTC 1040			
TGGGACGACTGTGTCAAATATAATTGCACGGTAGTGCACT 1080			
ACATCGGTGAAATATGCCGCTACCTGCTAAGGCAGCCGGT 1120			
TCGCGATGTAGAGCGGCGGCACCGCGTGCGCCTGGCCGTG 1160			
GGTAACGGA CTGCGGCCAGCCATCTGGGAGGAGTTCACGC 1200			

FIG. 56A

1210 1220 1230 1240

AGGGTTTTCGGTGTGCGACAGATTGGCGAGTTCTACGGCGC 1240  
CACCGAATGCAACTGCAGCATTGCCAACATGGACGGCAAG 1280  
GTCGGCTCCTGCGGCTTCAACAGCCGTATCCTCACGCATG 1320  
TGTACCCCATCCGTCTGGTCAAGGTCAACGAGGACACGAT 1360  
GGAGCCACTGAGGGACTCCCAAGGCCCTCTGCATCCCGTGC 1400

1410 1420 1430 1440

CAGCCCGGGGAACCTGGGCTTCTCGTGGGCCAGATCAACC 1440  
AGCAAGACCCTCTGCGGCGCTTCGATGGCTATGTTAGTGA 1480  
CAGCGCCACCAACAAGAAGATTGCCACAGCGTGTTCGA 1520  
AAGGGGGACAGCGCCTACCTTTCAGGTGACGTGCTAGTGA 1560  
TGGACGAGCTGGGGTACATGTACTTCCGTGACCGCAGCGG 1600

1610 1620 1630 1640

GGATACCTTCCGATGGCGCGGCGAGAACGTATCCACCACG 1640  
GAGGTGGAAGCCGTGCTGAGCCGCCTGTTGGGCCAGACGG 1680  
ACGTGGCTGTGTATGGAGTGGCTGTGCCAGGAGTGGAGGG 1720  
GAAAAGCGGCATGGCGGCCATTGCAGACCCCCACAACCAG 1760  
CTGGACCCTAACTCAATGTACCAGGAATTGCAGAAGGTTT 1800

1810 1820 1830 1840

TTGCATCCTATGCCCAGCCCATCTTCCTGCGTCTTCTGCC 1840  
CCAAGTGGATACAACAGGCACCTTCAAGATCCAGAAGACC 1880  
CGACTACAGCGTGAAGGCTTTGACCCCCGCCAGACCTCAG 1920  
ACCGGCTCTTCTTTCTAGACCTGAAACAGGGACGCTACCT 1960  
ACCCCTGGATGAGAGAGTCCATGCCCGCATCTGCGCAGGC 2000

2010 2020 2030 2040

GACTTCTCACTCTGAGCCTGGTGGTGGGATGGCCCTGGA 2040  
CTTGTGAGACCAGGGAGCCGGACACCCCTGTTCAAGGTGTT 2080  
TCTCCTGCTTGGCCACGTGCGCCAGCAGCACCTGTGGGTGC 2120  
AGGAAACTGGAACCTGAGTGGCCGGGTGTCCCTTTCTTAC 2160  
AACCACCATGCACACATCTAGCCTCTGCCTTGGTCTTTT 2200

2210	2220	2230	2240
TCTCCATCTCTTTCCCTCCGTGCCCAGCAGGAGCCCCACAG	2240		
ACACATTGGCTGCTGTGTCTGCAGTGGGACCGGTGTCTA	2280		
GGGGTCCATGCTGCAGGCTGTGACCCGCACTGGTGCCAC	2320		
CTCCCTTCCCCATTGTGCCTTAGGTTCTCCACTGTGCGC	2360		
CGGTGAAGCAAGTGGGGACCCACATAGCTGTTGTCCCTGC	2400		
2410	2420	2430	2440
TGAGGGTTGGTAGCAAATGCACCCTCATGTCAGCTGGGAG	2440		
ACACATGCAGTCTCCCACTGACCCCCAATCAACTGAAGAT	2480		
ACTGTTTTGTATTATTGTTTTGAGATAGGGTCTCACTGTG	2520		
GAGGCCAAGCTGGCCTCAGGCTCACCCTCTACTGCCTCC	2560		
GGGCACCAGCCTGCAGTTTGATGACATGTATGCACTATTG	2600		

FIG. 56C

2610	2620	2630	2640
TTCTAAGGGTCTTCTGAGTCCCTGCTTTCCCCTCATGTCC	2640		
TAAACCTTCCAGAACTGACTCTGATCACTTGGATGTAGC	2680		
TAGTGTTGGCCCTGCCACGTGTGTCAATTCAGGGGTCCC	2720		
CAGGCATCATCTCTGGAGGCCCTAACCTTGGCAAAGCTTG	2760		
GATGTCCTCACATCACAGCAGGAGACCCAGGAAGGTTGCT	2800		
2810	2820	2830	2840
GTGGTGTCTCTTGGGCACCCCTGGCGGCAGCCGTGGACAT	2840		
GCTTCCCTGCTGTGATAGCCCAAACCTGTTGCCTATGACAT	2880		
TTGAGGTCTACCCTTCTGGCTGCCATGGTCCCCATTGAGA	2920		
TCTTTGGTGACTCACCTCAGCCACCAAGCCAGGCCTCTGC	2960		
CTTCCTTCAGCTCTAAGGGCATGAAGGGTGTGGACAGAGC	3000		
3010	3020	3030	3040
AGCCACAGGCTGCCCCACAGTCACCCACATGCAAGTGTAT	3040		
TTCCTTGTTTGTGTTTTAAAAAATAAACATGCTGAGCCTTG	3080		
AAAAAAAAAAAAAAAAAAAAA	3098		

FIG. 56D



10	20	30	40	
MRTPGAGTASVASLGLLWLLGLPWTWSAAAAFGVYVGSGG 40				
WRFLRIVCKTARRODLFGLSVLIRVRLELRRHRRAGDTIPR 80				
IFQAVAQRQPERLALVDASSGICWTFAQLDTYSNAVANLF 120				
LQLGFAPGDVVAVFLEGRPEFVGLWLGLAKAGVVAALLNV 160				
NLRREPLAFCLGTSAAKALIYGGEMAAVAEVSEQLGKSL 200				
210	220	230	240	
LKFCSGDLGPESVLPDTQLLDPMLAEAPTTPLAQAPGKGM 240				
DDRLFYIYTS GTTGLPKAAIVVHSRYR IAAFGHHSYSMR 280				
ANDVLYDCLPLYHSAGNIMGVGQCIIYGLTVVLRKKFSAS 320				
RFWDDCVKYNCTVVQYIGEICRYLLRQPV RDVERRHRVRL 360				
AVGNGLRPAIWEEFTQGFGVRQIGEFYGATECNC SIANMD 400				
410	420	430	440	
GKVGSCGFNSRILTHVYPIRLVKVNEDTMEPLRDSQGLCI 440				
PCQPGEPGLLVGQINQODPLRRFDGYVSDSATNKKIAHSV 480				
FRKGDSAYLSGDVLVMDELGYMYFRDRSGDTFRWRGENVS 520				
TTEVEAVLSRLLGQTDVAVYGVAVPGVEGKSGMAAIADPH 560				
NQLDPNSMYQELQKVLASYAQPIFLRLLPQVDTTGTFKIQ 600				
610	620	630	640	
KTRLQREGFDPROTS DRLFFLDLKGGRYLPLOERVHARIC 640				
AGDFSL 646				

FIG. 57

FIG. 58A

1210 1220 1230 1240  
ACAACATCCGTGCCAAGTCTCTGCTACACTGCTTTCAGTG 1240  
CTGCGGGGCGAAGGTGCTGCTGGCCTCCCCAGAGCTACAC 1280  
GAAGCTGTGCGAGGAGGTTCTTCCAACCCTGAAAAAGGAGG 1320  
GCGTGTCGCTTCTACGTAAGCAGAACTTCTAACACTAA 1360  
TGGCGTGGACACAGTACTGGACAAAGTAGACGGGGTGTCTG 1400  
1410 1420 1430 1440  
GCGGACCCCATCCCGGAGTCGTGGAGGTCTGAAGTCACGT 1440  
TCACCACACCCGCAGTCTACATATATACTTCGGGCACCAC 1480  
AGGTCTTCCAAAGGCTGCAACCATTAAATCACCATCGCCTC 1520  
TGGTATGGGACCAGCCTTGCCCTGAGGTCCGGAATTAAGG 1560  
CTCATGACGTCATCTACACCACCATGCCCCTGTACCACAG 1600  
1610 1620 1630 1640  
CGCGGCGCTCATGATTGGCCTCCACGGATGCATTGTGGTT 1640  
GGGGCTACATTTGCTTTGCGGAGCAAATTTTCAGCCAGCC 1680  
AGTTTTTGGGACGACTGCAGGAAATACAACGCCACTGTCTAT 1720  
TCAGTACATCGGTGAACTGCTTCGGTACCTCTGCAACACG 1760  
CCCCAGAAACCAAATGACCGGGACCACAAAGTGAAAATAG 1800  
1810 1820 1830 1840  
CACTAGGAAATGGCTTACGAGGAGATGTGTGGAGAGAGTT 1840  
CATCAAGAGATTTGGGGACATTCACATTTATGAGTTCTAC 1880  
GCTTCCACTGAAGGCAACATTGGATTTATGAACTATCCAA 1920  
GAAAAATCGGAGCTGTTGGAAGAGAAAATTACCTACAAAA 1960  
AAAAGTTGTAAGGCACGAGCTGATCAAGTATGACGTGGAG 2000  
2010 2020 2030 2040  
AAGGATGAGCCTGTCCGTGATGCAAATGGATATTGCATCA 2040  
AAGTCCCCAAAGGAGAGGTTGGACTCTTGATTTGCAAAAT 2080  
CACAGAGCTCACACCATTTTTTTGGCTATGCTGGAGGAAAG 2120  
ACCCAGACAGAGAAGAAAAAGCTCAGAGATGTTTTTAAGA 2160  
AAGGAGACGTCTACTTCAACAGTGGCGATCTCCTGATGAT 2200

FIG. 58B

2610 2620 2630 2640

GTGCCCATGACTGAGGACATTTATAATGCCATAATTGATA 2640  
AGACTCTGAAGCTCTGAATGTTGCCTGGCTCCTAACACTT 2680  
CCAGAAAGAAACACAATAGGCCTAGCATAGCCCCTTCACA 2720  
TGTGTAATCCAACTTTAACTTGATTAAAGGTTATAGGTGT 2760  
GATTTTTCTTAGGAAATTATTCATTTAAAGGACAATTGTT 2800

2810 2820 2830 2840

TGTTTGTGGTTTGTGGTTTATTAATTACACCAGAACGTT 2840  
TGCAAGTAAAAAGATTTAAAGTCACTTATTTTTCAATGTG 2880  
CACCTGCCATTTGTCTTGCAAACCTAGCTTCTTGGAGAG 2920  
AGGGCCTTATTTTTTTAAAGACATAATAAACTATGTAAAC 2960  
ACT 2963

FIG. 58D

	10	20	30	40	
MLPVLYTGLAGLLLLPLLLTCCCPYLLQDVRFFLQLANMA	40				
RQVRSYRQRRPVRTILHVFLEQARKTPHKPFLLFRDETLT	80				
YAQVDRRSNQVARALHDHLGLRQGDVALFMGNEPAYVWL	120				
WLGLLLKLGCPMACLNYNIRAKSLLHCFQCCGAKVLLASPE	160				
LHEAVEEVLP TLKKEGVS VFYVSRTSNTNGVDTVLDKVDG	200				
	210	220	230	240	
VSADPIPESWRSEVTFTTPAVYIYTS GTTG LPKAATINHH	240				
RLWYGTSLALRSGIKAH DVIYTTMPLYHSAALMIGLHGCI	280				
VVGATFALRSKFSASQFWDDCRKYNATVIQYIGELLRYLC	320				
NTPOKPNDRDHKVKIALGNGLRGDVWREFIKRFGDIHIYE	360				
FYASTEIGNIGFMNYP RKIGAVGRENYLQKKVVRHEL IKYD	400				
	410	420	430	440	
VEKDEPVRDANGYCIKVPKG EVGLLICKITELTPFFGYAG	440				
GKTQTEKKKLRDVFKKGDVYFNSGDLLMIDRENFIYFHDR	480				
VGDTFRWKGENVATTEVADI VGLVDFVEEVN VYGVPVPGH	520				
EGRIGMASIKMKENYEFNGKKLFQHI SEYLPSYSRPRFLR	560				
IQDTIEITGTFKHKVTLMEEGFNPSVIKDTLYFMDDTEK	600				
	610	620	630	640	
TYVPMTEDIYNAIIDKTLKL	620				

FIG. 59

FIG. 60A

1210            1220            1230            1240  
 \_\_\_\_\_  
 GGCGAGGAGAAGCTGTGATTTCCCCACATCCCTCTGAGG 1240  
 GCCAGAGGATGCTGGATTTCAGAGCCCCAGCTTCCACTCCA 1280  
 GAAGGGGTCTGGGCAAGGCCAGACCAAAGCTAGCAGGGCC 1320  
 CGCACCTTCACCCTAGGTGCTGATCCCCCT 1350

FIG. 60B

10            20            30            40  
 \_\_\_\_\_  
 DQLFYIYTS GTTGLPKAAIVVHSRYRMAALVYYGFRMRP 40  
 DDIVYDCLPLYHSAGNIVGIGQCVLHGMTVVIRKKFSASR 80  
 FWDDCIKYNCTIVQYIGELCRYLLNQPPREAESRHKVRMA 120  
 LGNGLRQSIWTD FSSRFHIPKVAEFYGATECNC SLGNFDS 160  
 QVGACGFNSRILSFVYPIRLVRVNEDTMELIRGPDGVCIP 200  
 210            220            230            240  
 \_\_\_\_\_  
 CQPGQPGQLVGR IIOQDPLRRFDGYLNQGANNKKIASDVF 240  
 KKGDAQAYLTGDVLVMDLGYLYFRDRTGDTFRWKGENYST 280  
 TEVEGTLSRLLQMAADVAYYGVEVPGAEGRAGMAAVASPTS 320  
 NCDLESFAQTLKKELPLYARPIFLRFLPELHKTGTFFKFQK 360  
 TELRKEGFDPSVVKDPLFYLDARTGCYVALDQEAYTRIQA 400  
 410            420            430            440  
 \_\_\_\_\_  
 GEEKL 405

FIG. 61





1210	1220	1230	1240
GGCAAGGTCGGCTCCTGCGGCTTCAACAGCCGTATCCTCA	1240		
CGCATGTGTACCCCATCCGTCTGGTCAAGGTCAATGAGGA	1280		
CACGATGGAGCCACTGCGGGACTCCGAGGGCCTCTGCATC	1320		
CCGTGCCAGCCCCGGGGAACCCGGCCTTCTCGTGCGGCCAGA	1360		
TCAACCAGCAGGACCCTCTGCGGCGTTTCGATGGTTATGT	1400		
1410	1420	1430	1440
TAGTGACAGTGCCACCAACAAGAAGATTGCCCCACAGCGTT	1440		
TTCCGAAAGGGCGATAGCGCCTACCTCTCAGGTGACGTGC	1480		
TAGTGATGGACGAGCTGGGCTACATGTATTTCCGTGACCG	1520		
CAGCGGGGACACCTTCCGCTGGCGCGGGGAGAACGTGTCC	1560		
ACCACGGAGGTGGAAGCCGTGCTGAGCCGCCTACTGGGCC	1600		
1610	1620	1630	1640
AGACGGACGTGGCTGTGTATGGGGTGGCTGTGCCAGGAGT	1640		
GGAGGGGAAAGCTGGCATGGCAGCCATCGCAGATCCCCAC	1680		
AGCCAGTTGGACCCTAACTCAATGTACCAGGAATTACAGA	1720		
AGGTTCTTGCATCCTATGCTCGGCCCATCTTCCCTGCGTCT	1760		
TCTGCCCCAGGTGGATACCACAGGCACCTTCAAGATCCAG	1800		
1810	1820	1830	1840
AAGACCCGGCTGCAGCGTGAAGGCTTTGACCCCGTCAGA	1840		
CCTCAGACAGGCTCTTCTTTCTAGACCTGAAGTCCGGCAC	1880		
GAGGTATCTACCCCTGGATGAGAGAGTCCATGCCCGCATT	1920		
TGCGCAGGCGACTTCTCACTCTGAGCCTGGAGAGTGGGCT	1960		
GGGCCTGGACTCCTGAGACCTGGGAGCCTGACACCCCTCT	2000		
2010	2020	2030	2040
TCGGGTGCTTCTCCTGCCTGGCCACATGGACAGCAGCACC	2040		
TGTGAGAGTAGGAAAATGGAACCTGAGTGGCTGGGACCCC	2080		
TCTCCTACTTCCCACTATGCATCCATTTTGCTCTGCCTT	2120		
GATCTTTTTTCTCCATCTCTTTTCTCCCTACCCAGCAGGAG	2160		
CCCCACAAACACATGTTGGCTGCTGTGTCTCTGCAGTTGA	2200		

FIG. 62B

2210 2220 2230 2240  
CCAGTGTCCAGGGGTACAGGCTTCAGGCTGTGACCCACAC 2240  
TGGTACCCACCTCCCTTTTCCTATTTTGCCTTAGGTTTCATC 2280  
CACGGTTCCCCTGTGGAGCAAGTGGGGGCCCACATAGCTG 2320  
CTGTCCCTGCTGAGGGTTGGTAGCAATCACACCCTCATGT 2360  
CAGCTGGGAGACACGCGCAGTCTCCCCTGACCCCCAATC 2400  
2410 2420 2430 2440  
AACTGAAAATATTGTTTTGACTACTTTTTGTTTTTTTGT 2440  
TTTTTGTTTTTTTTTTTTTCGAGACAGAGTTTCTCTGTA 2480  
TAGCCCTGGCTGTCCTGGAACCTCACTTTGTAGACCAGGCT 2520  
GGCCTCGAACTCAAAAATCCTCCTGACTCTGCCTCTGCTT 2560  
CCCAAGTGCTGGGATTAAAGACGTGCGCCACCACCGCCTG 2600

FIG. 62C

2610 2620 2630 2640  
GCTGTTTTGTATTTTTGTTTTGTTTTGACGATAGGGTCTC 2640  
ACTGTGGAGGCCAAGCTGGCCTCAGACTCCCCACCCCAT 2680  
GCCTCTGGGCACCATTTCTATATTCTCAGACTGATGACAAT 2720  
GCACTAGTGTCCCTAGGAGTCTTGAGTCTGCACTTTCCCC 2760  
TCATAGCCTCAAGCTTCCAGAACTGACTCTGATCACTTGG 2800  
2810 2820 2830 2840  
ATGTGGCTAGTGTTGGCTCTACCCACATGTGTCAATTCAG 2840  
GGGTCCCCAGGCATAGTCTCTGGAAGCCCTCACCCGGA 2880  
AAGCTTGAGAGAGACCCAGGAAGGTTGTTGTGTTCTCTTGG 2920  
GCACCCCTGGTGGCAGTCCTGGGCATGCTTCCGCACTGT 2960  
ACTGGTGCATATAGCCCAGACCTATGACATTTGAGGTCTA 3000  
3010 3020 3030 3040  
CCCTTCTGGCTCCTGTGGTCCCCATTGAGATCCTTGGTGA 3040  
CTCACCTCAGTCACCAAGCAGAGCCTCTGCCTGCCTTCAT 3080  
CTTCAAGGTCATGAAGGATGTGGACAGAGCAGCTACAGGC 3120  
TGCCAGCAGTCAACCACATGAGAGTGTTACTTCCTTGTTG 3160  
GTTTTTAAAAAATAAATGTGCTGAGCCTCGAAAAAAAAA 3200  
3210 3220 3230 3240  
AAAAAAAAAAAAAAAAA 3217

FIG. 62D

**SUBSTITUTE SHEET (RULE 26)**

FIG. 64A

FIG. 64B

10	20	30	40	
<hr/>				
MLPVLYTGLAGLLLLPLLLTCCCPYLLQDVRYFLRLANMA	40			
RRVRSYRQRRPVRTILRAFLEQARKTPHKPFLLFRDETLT	80			
YAQVDRRSNQVARALHDQLGLRQGDCVALFMGNEPAYVWI	120			
WLGLLLKLGCPMACLNYNIRAKSLLHCFQCCGAKVLLASPD	160			
LQEAVEEVLP TLKKDAVS VFYVSRTSNTNGVD TILDKVDG	200			
210	220	230	240	
<hr/>				
VSAEPTPESWRSEVTFTTPAVYIYTS GTTGLPKSGT INHH	240			
RLRYGTSLAMSSGNHGQGCHLYQQCPCSNSATLKIGLHGC	280			
ILGWGYFNLGGANSQASQFWERLAGNTTSTVIQYIGELLR	320			
YLCNTPQKPNDRDHKVKKALGNGLRGDVWREFIKRFGDIH	360			
VYEFYASTE GNIGFVNYP RKIGAVGRANYLQRKVARYELI	400			
410	420	430	440	
<hr/>				
KYDVEKDEPVRDANGYCIKVPKGEVGLLVCKITQLTPFIG	440			
YAGGKTQTEKKKLRDVFKKGDIYFNSGDLLMIDRENFVYF	480			
HDRVGDTRFWKGENVATTEVADI VGLVDFVEEVN VYGVPV	520			
PGHEGRIGMASLKIKENYEFNGKKLFQHIAEYLPSYARPR	560			
FLRIODTIEITGTFKHKRVTLMEEGFNPTVIKDTLYFMDD	600			
610	620	630	640	
<hr/>				
AEKTFVPMTENIYNAIDKTLKL	623			

FIG. 65

FIG. 66A



1210 1220 1230 1240  
GGGCACTGCATGACCACATCTCCAGGTGAGCCAGGCCTAC 1240  
TGGTGGCCCCAGTGAGCCAGCAGTCCCCCTTCCTGGGCTA 1280  
TGCTGGGGCTCCGGAGCTGGCCAAGGACAAGCTGCTGAAG 1320  
GATGTCTTCTGGTCTGGGGACGTTTTCTTCAATACTGGGG 1360  
ACCTCTTGGTCTGTGATGAGCAAGGCTTCTTCACTTCCA 1400  
1410 1420 1430 1440  
CGATCGTACTGGAGACACCATCAGGTGGAAGGGAGAGAAT 1440  
GTGGCCACAACCTGAAGTGGCTGAGGTCTTGGAGACCCTGG 1480  
ACTTCCTTCAGGAGGTGAACATCTATGGAGTCACGGTGCC 1520  
AGGGCACGAAGGCAGGGCAGGCATGGCGGCCTTGGCTCTG 1560  
CGGCCCCCGCAGGCTCTGAACCTGGTGCAGCTCTACAGCC 1600  
1610 1620 1630 1640  
ATGTTTCTGAGAACTTGCCACCGTATGCCCCGACCTCGGTT 1640  
TCTCAGGCTCCAGGAATCTTTGGCCACTACTGAGACCTTC 1680  
AAACAGCAGAAGGTTAGGATGGCCAATGAGGGCTTTGACC 1720  
CCAGTGTA CTGTCTGACCCACTCTATGTTCTGGACCAAGA 1760  
TATAGGGGCCTACCTGCCCCCTCACACCTGCCCCGTACAGT 1800  
1810 1820 1830 1840  
GCCCTCCTGTCTGGAGACCTTTCGAATCTGAAACCTTCCAC 1840  
TTGAGGGGAGGGGCTCGGAGGGTACAGGCCACCATGGCTGC 1880  
ACCAGGGAGGGTTTTTCGGGTATCTTTTGTATATGGAGTCA 1920  
TTATTTTGTAAATAAACAGCTGGAGCTTAAAAAAAAAAAAAA 1960  
AA 1998

FIG. 66B

10	20	30	40	
<hr/>				
ESSESGCSLAWRLAYLAREQPTHTFLIHGAQRFSYAEAE	40			
ESNRIARAFLRARGWTGGRRGSGRGSTEEGARVAPPAGDA	80			
AARGTTAPPLAPGATVALLLPAGPDFLWIWFGGLAKAGLRT	120			
AFVPTALRRGPLLHCLRSCGASALVLATEFLESLEPDLP	160			
LRAMGLHLWATGPETNVAGISNLLSEAADQVDEPVPGYLS	200			
210	220	230	240	
<hr/>				
APQNI	MDTCLYIFTSGTTGLPKAARISHLKVLCQGFYHL	240		
CGVHQEDVIYLALPLYHMSGSLLGIVGCLGIGATVVLKPK	280			
FSASQFWDDCQKHRVTVFQYIGELCRYLVNQPPSKAEFDH	320			
KVRLAVGSGLRPDTWERFLRRFGPLQILETYGMTEGNVAT	360			
FNYTGROGAVGRASWLYKHIFPFSLIRYDVMTGEP	400			
410	420	430	440	
<hr/>				
GHCM	TTSPGEPGLLVAPVSQOSPFLGYAGAPELAKDKLLK	440		
DVFW	SGOVFFNTGOLLVCDEQGFLHFHRTGDTIRWKGEN	480		
VATTE	VAEVLETLOFLQEVNIYGVTPGHEGRAGMAALAL	520		
RPPQ	ALNLVQLYSHVSENLPYARPRFLRLQESLATTETF	560		
KQKQ	VRMANEGFDPSVLSOPLYVLDQDIGAYLPLTPARYS	600		
610	620	630	640	
<hr/>				
ALLSGDLRI	609			

FIG. 67

10 20 30 40

ATGCTGCTTGGAGCCTCTCTGGTGGGGGCGCTACTGTTCT 40  
CCAAGCTAGTGCTGAAGCTGCCCTGGACCCAGGTGGGATT 80  
CTCCCTGTTGCTCCTGTACTTGGGGTCTGGTGGCTGGCGT 120  
TTCATCCGGGTCTTTCATCAAGACGGTCAGGAGAGATATCT 160  
TTGGTGGCATGGTGCTCCTGAAGGTGAAGACCAAGGTGCG 200

210 220 230 240

ACGGTACCTTCAGGAGCGGAAGACGGTGCCCCTGCTGTTT 240  
GCTTCAATGGTACAGCGCCACCCGGACAAGACAGCCCTGA 280  
TTTTTCGAGGGCACAGACACTCACTGGACCTTCCGCCAGCT 320  
GGATGAGTACTCCAGTAGTGTGGCCAACTTCTGCAGGCC 360  
CGGGGCCTGGCCTCAGGCAATGTAGTTGCCCTCTTTATGG 400

410 420 430 440

AAAACCGCAATGAGTTTGTGGGTCTGTGGCTAGGCATGGC 440  
CAAGCTEGGGCGTGGAGGCGGCTCTCATCAACACCAACCTT 480  
AGGCGGGGATGCCCTGCGCCACTGTCTTGACACCTCAAAGG 520  
CACGAGCTCTCATCTTTGGCAGTGAGATGGCCTCAGCTAT 560  
CTGTGAGATCCATGCTAGCCTGGAGCCCACACTCAGCCTC 600

610 620 630 640

TTCTGCTCTGGATCCTGGGAGCCCAGCACAGTGCCCGTCA 640  
GCACAGAGCATCTGGACCCTCTTCTGGAAGATGCCCCGAA 680  
GCACCTGCCCAGTCACCCAGACAAGGGTTTTACAGATAAG 720  
CTCTTCTACATCTACACATCGGGCACCACGGGGCTACCCA 760  
AAGCTGCCATTGTGGTGCACAGCAGGTATTATCGTATGGC 800

810 820 830 840

TTCCCTGGTGTACTATGGATTCCGCATGCGGCCTGATGAC 840  
ATTGTCTATGACTGCCTCCCCCTCTACCACTCAAGCAGGA 880  
AACATCGTGGGGATTGGCAGTGCTTACTCCACGGCATGAC 920  
TGTGGTGATCCGGAAGAAGTTCTCAGCCTCCCGGTTCTGG 960  
GATGATTGTATCAAGTACAACCTGCACAGTGGTACAGTACA 1000

1010 1020 1030 1040

TTGGCGAGCTCTGCCGCTACCTCCTGAACCAGCCACCCCG 1040  
TGAGGCTGAGTCTCGGCACAAGGTGCGCATGGCACTGGGC 1080  
AACGGTCTCCGGCAGTCCATCTGGACCGACTTCTCCAGCC 1120  
GTTTCCACATCCCCCAGGTGGCTGAGTTCTATGGGGCCAC 1160  
TGAATGCAACTGTAGCCTGGGCAACTTTGACAGCCGGGTG 1200

FIG. 68A

1210 1220 1230 1240  
GGGGCCTGTGGCTTCAATAGCCGCATCCTGTCCTTTGTGT 1240  
ACCCTATCCGTTTGGTACGTGTCAATGAGGATACCATGGA 1280  
ACTGATCCGGGGACCCGATGGAGTCTGCATTCCCTGTCAA 1320  
CCAGGTCAGCCAGGCCAGCTGGTGGGTGCGCATCATCCAGC 1360  
AGGACCCTCTGCGCCGTTTCGACGGGTACCTCAACCAGGG 1400  
1410 1420 1430 1440  
TGCCAACAACAAGAAGATTGCTAATGATGTCTTCAAGAAG 1440  
GGGGACCAAGCCTACCTCACTGGTGACGTCCTGGTGATGG 1480  
ATGAGCTGGGTTACCTGTACTTCCGAGATCGCACTGGGGGA 1520  
CACGTTCCGCTGGAAAGGGGAGAATGTATCTACCACTGAG 1560  
GTGGAGGGCACACTCAGCCGCCTGCTTCATATGGCAGATG 1600  
1610 1620 1630 1640  
TGGCAGTTTATGGTGTTGAGGTGCCAGGAAGTGAAGGCCG 1640  
AGCAGGAATGGCTGCCGTTGCAAGTCCCATCAGCAACTGT 1680  
GACCTGGAGAGCTTTGCACAGACCTTGAAAAAGGAGCTGC 1720  
CTCTGTATGCCCCGCCCCATCTTCCTGCGCTTCTTGCCTGA 1760  
GCTGCACAAGACAGGGACCTTCAAGTTCCAGAAGACAGAG 1800  
1810 1820 1830 1840  
TTGCGGAAGGAGGGCTTTGACCCATCTGTTGTGAAAGACC 1840  
CGCTGTTCTATCTGGATGCTCGGAAGGGCTGCTACGTTGC 1880  
ACTGGACCAGGAGGCCTATACCCGCATCCAGGCAGGCCAG 1920  
GAGAAGCTGTGATTTCCCCCTACATCCCTCTGAGGGCCAG 1960  
AAGATGCTGGATTTCAGAGCCCTAGCGTCCACCCCAGAGGG 2000

FIG. 68B

2010                      2020                      20  
 TCCTGGGCAATGCCAGACCAAAGCTAGCAGGGCCCGCACC 2040  
 TCCGCCCCCTAGGTGCTGATCTCCCCTCTCCCAAAGTGGCA 2080  
 AGTGACTCACTGCCGCTTCCCCGACCCTCCAGAGGCTTTC 2120  
 TGTGAAAGTCTCATCCAAGCTGTGTCTTCTGGTCCAGGCG 2160  
 TGGCCCCCTGGCCCCAGGGTTTCTGATAGGCTCCTTTAGGA 2200  
 2210                      2220                      2230                      2240  
 TGGTATCTTGGGTCCAGCGGGCCAGGGTGTGGGAGAGGAG 2240  
 TCACTAAGATCCCTCCAATCAGAAGGGAGCTTACAAAGGA 2280  
 ACCAAGGCAAAGCCTGTAGACTCAGGAAGCTAAGTGGCCA 2320  
 GAGACTATAGTGGCCAGTCATCCCATGTCCACAGAGGATC 2360  
 TTGGTCCAGAGCTGCCAAAGTGTACCTCTCCCTGCCTGC 2400  
 2410                      2420                      2430                      2440  
 ACCTCTGGGGAAAAGAGGACAGCATGTGGCCACTGGGCAC 2440  
 CTGTCTCAAGAAAGTCAGGATCACACACTCAGTCCTTGTTT 2480  
 CTCCAGGTTCCCTTGTTCTTGTTCTCGGGGAGGGAGGGACG 2520  
 AGTGTCCTGTCTGTCCTTCCTGCCTGTCTGTGAGTCTGTG 2560  
 TTGCTTCTCCATCTGTCCTAGCCTGAGTGTGGGTGGAACA 2600

FIG. 68C

2610                      2620                      2630                      2640  
 GGCATGAGGAGAGTGTGGCTCAGGGGCCAATAAACTCTGC 2640  
 CTTGACTCCTCTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2680  
 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2710

FIG. 68D

10	20	30	40
<hr/>			
MLLGASLVGALLFSKLVCLKPWTQVGFSLLLLYLGSGGWR	40		
FIRVFIKTVRRDIFGGMVLLKVKTkVRRYLOERKTVPLLF	80		
ASMVQRHPDKTALIFEGTDTHWTFRQLDEYSSSVANFLQA	120		
RGLASGNVVALFMENRNEFVGLWLGMALGVEAALINTNL	160		
RRDALRHCLDTSKARALIFGSEMASAICEIHASLEPTLSL	200		
210	220	230	240
<hr/>			
FCSGSWEPSTVPVSTEHLDPILLEDAPKHLPSHPDKGFTDK	240		
LFYIYTSGTTGLPKAAIVVHSRYRMAASLVYYGFRMRPDD	280		
IVYDCLPLYHSSRKHRGDWQCLLHGMTVVIRKKFSASRFW	320		
DDCIKYNCTVVQYIGELCRYLLNQPPREAESRHKVRMALG	360		
NGLRQSIWTDFFSSRFHIPOVAEFYGATECNCSLGNFDSRV	400		
410	420	430	440
<hr/>			
GACGFNSRILSFVYPIRLVRVNEDTMELIRGPDGVCIPCQ	440		
PGOPGQLVGRIIQQDPLRRFDGYLNQGANNNKIANDVFKK	480		
GDQAYLTGDVLVMDDELGYLYFRDRTGDTFRWKGENVSTTE	520		
VEGTLSRLLHMADVAVYGVVPGTEGRAGMAAVASPI SNC	560		
DLESFAOTLKKELPLYARPIFLRFLPELHKTGTGFKFQKTE	600		
610	620	630	640
<hr/>			
LRKEGFDPSPVVKDPLFYLDARKGCYVALDQEAAYTRIQAGE	640		
EKL	643		

FIG. 69

10	20	30	40	
<hr/>				
CACTCATCAGAGCTAAGAGAGACTACACGCTCTCATCTAC	40			
TTCAGAAAGAGCCAATGCCATGGGTATTTGGAAGAACTA	80			
ACCTTACTGCTGTTGCTGCTTCTGCTGGTTGGCCTGGGGC	120			
AGCCCCCATGGCCAGCAGCTATGGCTCTGGCCCTGCGTTG	160			
GTTCTCTGGGAGACCCACATGCCTTGTGCTGCTTGGCTTG	200			
210	220	230	240	
<hr/>				
GCATTGCTGGGCAGACCCTGGATCAGCTCCTGGATGCCCC	240			
ACTGGCTGAGCCTGGTAGGAGCAGCTCTTACCTTATTCCT	280			
ATTGCCTCTACAGCCACCCCCAGGGCTACGCTGGCTGCAT	320			
AAAGATGTGGCTTTTACCTTCAAGATGCTTTTCTATGGCC	360			
TAAAGTTCAGGCGACGCCTTAACAACATCCTCCAGAGAC	400			
410	420	430	440	
<hr/>				
CTTTGTGGATGCTTTAGAGCGGCAAGCACTGGCATGGCCT	440			
GACCGGGTGGCCTTGGTGTGTACTGGGTCTGAGGGCTCCT	480			
CAATCACAAATAGCCAGCTGGATGCCAGGTCCTGTCAGGC	520			
AGCATGGGTCTTGAAAGCAAAGCTGAAGGATGCCGTAATC	560			
CAGAACACAAGAGATGCTGCTGCTATCTTAGTTCTCCCGT	600			
610	620	630	640	
<hr/>				
CCAAGACCATTTCTGCTTTGAGTGTGTTTCTGGGGTTGGC	640			
CAAGTTGGGCTGCCCTGTGGCCTGGATCAATCCACACAGC	680			
CGAGGGATGCCCTTGCTACACTCTGTACGGAGCTCTGGGG	720			
CCAGTGTGCTGATTGTGGATCCAGACCTCCAGGAGAACCT	760			
GGAAGAAGTCCTTCCCAAGCTGCTAGCTGAGAACATTAC	800			
810	820	830	840	
<hr/>				
TGCTTCTACCTTGGCCACAGCTCACCCACCCCGGGAGTAG	840			
AGGCTCTGGGAGCTTCCCTGGATGCTGCACCTTCTGACCC	880			
AGTACCTGCCAGCCTTCGAGCTACGATTAAGTGGAAATCT	920			
CCTGCCATATTCATCTTTACTTCAGGGACCACTGGACTCC	960			
CAAAGCCAGCCATCTTATCACATGAGCGGGTCATACAAGT	1000			
1010	1020	1030	1040	
<hr/>				
GAGCAACGTGCTGTCCTTCTGTGGATGCAGAGCTGATGAT	1040			
GTGGTCTATGACGTCCTACCTCTGTACCATACGATAGGGC	1080			
TTGTCCTTGGATTCCCTTGGCTGCTTACAAGTTGGAGCCAC	1120			
CTGTGTCCTGGCCCCCAAGTTCTCTGCCTCCCGATTCTGG	1160			
GCTGAGTGCCGGCAGCATGGCGTAACAGTGATCTTGTATG	1200			

FIG. 70A

1210	1220	1230	1240
TGGGTGAAATCCTGCGGTACTTGTGTAACGTCCCTGAGCA 1240			
ACCAGAAGACAAGATACATACAGTGCGCTTGGCCATGGGA 1280			
ACTGGACTTCGGGCAAATGTGTGGAAAACTTCCAGCAAC 1320			
GCTTTGGTCCCATTTCGGATCTGGGAATTCTACGGATCCAC 1360			
AGAGGGCAATGTGGGCTTAATGAACTATGTGGGCCACTGC 1400			
1410	1420	1430	1440
GGGGCTGTGGGAAGGACCAGCTGCATCCTTCGAATGCTGA 1440			
CTCCCTTTGAGCTTGTACAGTTCGACATAGAGACAGCAGA 1480			
GCCTCTGAGGGACAAACAGGGTTTTTGCATTCTGTGGAG 1520			
CCAGGAAAGCCAGGACTTCTTTTGGACCAAGGTTTCGAAAGA 1560			
ACCAACCCTTCCTGGGCTACCGTGGTTCCAGGCCGAGTC 1600			
1610	1620	1630	1640
CAATCGGAAACTTGTTCGAATGTACGACGCGTAGGAGAC 1640			
CTGTACTTCAACACTGGGGACGTGCTGACCTTGGACCAGG 1680			
AAGGCTTCTTCTACTTTCAAGACCGCCTTGGTGACACCTT 1720			
CCGGTGGAAAGGGCGAAAACGTATCTACTGGAGAGGTGGAG 1760			
TGTGTTTTGTCTAGCCTAGACTTCCTAGAGGAAGTCAATG 1800			
1810	1820	1830	1840
TCTATGGTGTGCCTGTGCCAGGGTGTGAGGGTAAGGTTGG 1840			
CATGGCTGCTGTGAAACTGGCTCCTGGGAAGACTTTTGAT 1880			
GGGCAGAAGCTATACCAGCATGTCCGCTCCTGGCTCCCTG 1920			
CCTATGCCACACCTCATTTTCATCCGTATCCAGGATTCCCT 1960			
GGAGATCACAAACACCTACAAGCTGGTAAAGTCACGGCTG 2000			
2010	2020	2030	2040
GTGCGTGAGGGTTTTGATGTGGGGATCATTGCTGACCCCC 2040			
TCTACATACTGGACAACAAGGCCAGACCTTCCGGAGTCT 2080			
GATGCCAGATGTGTACCAGGCTGTGTGTGAAGGAACCTGG 2120			
AATCTCTGACCACCTAGCCAACTGGAAGGCAATCCAAAAG 2160			
TGTAGAGATTGACACTAGTCAGCTTCACAAAGTTGTCCGG 2200			
2210	2220	2230	2240
GTTCCAGATGCCCATGGCCCAGTAGTACTTAGAGAATAAA 2240			
CTTGAATGTGTATACAAAAA			

FIG. 70B



10	20	30	40
..... ..... ..... .....			
MGIWKKL	TLLLLLLLL	VGLGQPPWPAAMALALRWFLGDPT	40
CLVLLGLALLGRPWISSWMPHWLSLVGAALTFLFLLPLQPP	80		
PGLRWLHKDVAFTFKMLFYGLKFRRRLNKHPPETFVDALE	120		
RQALAWPDRVALVCTGSEGSSITNSQLDARSCQAAWVLKA	160		
KLKDAV IQNTRDAAA ILVLP SKTISALSVFLGLAKLGCPV	200		
210	220	230	240
..... ..... ..... .....			
AWINPHSRGMPLLLHSVRSSGASVLIVDPDLQENLEEVLPK	240		
LLAENIHCFYLGHSSPTPGVEALGASLDAAPSDPVPASLR	280		
ATIKWKSPAIFIFTSGTTGLPKPAILSHERV IQVSNVLSF	320		
CGCRADDVVDVLP LYHTIGLVLGFLGCLQVGATCVLAPK	360		
FSASRFWAECRQHGVTVILYVGEILRYLCNVPEQPEDKIH	400		
410	420	430	440
..... ..... ..... .....			
TVRLAMGTGLRANVWKNFQQRFGPIRIWEFYGSTEGNVGL	440		
MNYVGHCGAVGRTSCILRMLTPFELVQFDIETAELRDKQ	480		
GFCIPVEPGKPGLLLTKVRKNQPF LGYRGSQAESNRKLVA	520		
NVRRVGDLYFNTGDVLTLDQEGFFYFQDR LGDTFRWKGEN	560		
VSTGEVECVLSSLD FLEEVNVYGVVPVPGCEGKVGMAAVKL	600		
610	620	630	640
..... ..... ..... .....			
APGKTFDGQKLYOHVRSWLPAYATPHFIRIQDSLEITNTY	640		
KLVKSR LVREGFDVGIIADPLYILDNKAQTFRSLMPDVYQ	680		
AVCEGTWNL	689		

FIG. 71

FIG. 72A

1210	1220	1230	1240
AAAGGAAATCCTTCTCGCGAATTCCTCGGATACGTCGATG 1240			
AAAAGGCCTCCGCGAAGAAGATTGTTAAGGATGTGTTCAA 1280			
GCATGGCGATATGGCTTTCATCTCCGGAGATCTGCTGGTT 1320			
GCCGACGAGAAGGGTTATCTGTACTTCAAGGATCGCACCG 1360			
GTGACACCTTCCGCTGGAAGGGCGAGAATGTTTCCACCAG 1400			
1410	1420	1430	1440
CGAGGTGGAGGCGCAAGTCAGCAATGTGGCCGGTTACAAG 1440			
GATACCGTCGTTTACGGCGTAACCATTCGCGACACCGAGG 1480			
GAAGGGCCGGCATGGCCGCCATCTATGATCCGGAGCGAGA 1520			
ATTGGACCTCGACGCTTTCGCCGCTAGCTTGGCCAAGGTG 1560			
CTGCCCCGCGTACGCTCGTCCCCAGATCATTCCGATTGCTCA 1600			
1610	1620	1630	1640
CCAAGGTGGACCTGACTGGAACCTTTAAGCTGCGCAAGGT 1640			
AGACCTGCAGAAGGAGGGCTACGATCCGAACGCGATCAAG 1680			
GACGCGCTGTACTACCAGACTTCCAAGGGTCGGTACGAGC 1720			
TGCTCACGCCCCAGGTTTACGACCAGGTGCAGCGCAACGA 1760			
AATCCGCTTCTAAGAGCTGCAATAGAGTTGTGTCTGAACC 1800			
1810	1820	1830	1840
TTGCCTTTTGCCCAATATGCTGTTAATTAGTTTGTAAAGGC 1840			
TAAGTGTAAGTAGAGGAAAATCGGGGAAAATCGGCAGCAAA 1880			
GATCATTCAGCCTAGGAGAGATGCATCCGAAGCACATTTT 1920			
CATGTCAACAATGCACTTTTGTATATCGTAAGCATATATA 1960			
TATCGTATATCGTAAACGTAGTTGTATCTGCATTTGTGTA 2000			
2010	2020	2030	2040
GATGATAGCCTCCTATACGCATTTCAATTGTTTTTAGCGT 2040			
GCTAAAGAACCTTGTTAAATGCAATTTTACGCTATTGTTTA 2080			
GTCAGTTTTTAGTGGCATTTACACTTCCATTCTCGTTGCGT 2120			
TTCGTTTTTTGCCTGTACATATGAGAAGCTCTGATGTTTTT 2160			
GTATCAAATAAAGTTTTTTTCTTACCACGGACCACGTGA 2200			
2210	2220	2230	2240
AAAAAAAAAAAAAAAAAAAAA 2221			

FIG. 72B

10 20 30 40  
 ALWAYIKLLRYTKRHERLNYTVADVFERNVQAHPDKVAVV 40  
 SETQRWTFRQVNEHANKVANVLQAQGYKKGDVVALLLENR 80  
 AEYVATWLGLSKIGVITPLINTNLRGPSLLHSITVAHCSA 120  
 LIYGEDFLEAVTDVAKDLPANLTLFQFNNENNNSETEKNI 160  
 POAKNLNALLTTASYEKPKNKTQVNHHDKLVIYITSGTTGL 200  
 210 220 230 240  
 PKAAVISHSRYLFIAAGIHYTMGFQEEDIFYTPLPLYHTA 240  
 GGIMCMGQSVLFGSTVSIKKFSASNYFADCAKYNATIGQ 280  
 YIGEMARYILATKPSEYDQKHRVRLVFGNGLRPQIWQFV 320  
 QRFNIAKVGEFYGATEGNANIMNHDNTVGAIGFVSRILPK 360  
 IYPISIIRADPDTGEPIRDNRGLCOLCAPNEPGVFIGKIV 400  
 410 420 430 440  
 KGNPSREFLGYYDEKASAKKIVKDVFKHGDMAFISGDLLV 440  
 ADEKGYLYFKDRTGDTFRWKGENVSTSEVEAQVSNVAGYK 480  
 DTVVYGVTIPTHEGRAGMAAIYDPERELDLDFVFAASLAKV 520  
 LPAYARPOIIRLLTKVDLTGTFKLRKVDLOKEGYDPNAIK 560  
 DALYYQTSKGRYELLTPQVYDQVORNEIRF 590

FIG. 73

10 20 30 40  
 AGTG TAGATACCA CAGGAACGTTTAAAATCCAGAAGACCA 40  
 GACTGCAAAGGGAAGGATACGATCCACGGCTCACA ACTGA 80  
 CCAGATCTACTTCCTAAACTCCAGAGCAGGGCGTTACGAG 120  
 CTTGTCAACGAGGAGCTGTACAATGCATTTGAACAAGGGC 160  
 AGGATTTCCCTTT 173

FIG. 74

10 20 30 40  
 SVDTTGTGFKIQKTRLQREGYDPRLTTDQIYFLNSRAGRYE 40  
 LVNEELYNAFEQGQDFP 57

FIG. 75

10 20 30 40  
ATGAAGCTGGAGGAGCTTGTGACAGTTATGCTTCTCACAG 40  
TGGCTGTCATTGCTCAGAATCTTCCGATTGGAGTAATATT 80  
GGCTGGAGTTCTTATTTTATACATCACAGTGGTTTCATGGA 120  
GATTTTCATTTATAGAAGTTATCTTACGTTGAATAGGGATT 160  
TAACAGGATTGGCTCTAATTATTGAAGTCAAAATCGACCT 200  
210 220 230 240  
ATGGTGGAGGTTGCATCAGAATAAAGGAATCCATGAACTG 240  
TTTTTGGATATTGTGAAAAAGAATCCAAATAAGCCGGCGA 280  
TGATTGACATCGAGACGAATACAACAGAAACATACGCAGA 320  
GTTCAATGCACATTGTAATAGATATGCCAATTATTTCCAG 360  
GGTCTTGGCTATCGATCCGGAGACGTTGTCGCCTTGTACA 400  
410 420 430 440  
TGGAGAACTCGGTTCGAGTTTGTGGCCGCGTGGATGGGACT 440  
CGCAAAAATCGGAGTTGTAACGGCTTGGATCAACTCGAAT 480  
TTGAAAAGAGAGCAACTTGTTTCATTGTATCACTGCGAGCA 520  
AGACAAAGGCGATTATCACAAGTGTAACACTTCAGAATAT 560  
TATGCTTGATGCTATCGATCAGAAGCTGTTTGATGTTGAG 600  
610 620 630 640  
GGAATTGAGGTTTACTCTGTCTGGAGAGCCCAAGAAGAATT 640  
CTGGATTCAAGAATCTCAAGAAGAAGTTGGATGCTCAAAT 680  
TACTACGGAACCAAGACCCCTTGACATAGTAGATTTTAAA 720  
AGTATTCTTTGCTTCATCTATACAAGTGGTACTACTGGAA 760  
TGCCAAAAGCCGCTGTTCATGAAGCACTTCAGATATTACTC 800  
810 820 830 840  
GATTGCCGTTGGAGCCGCAAAATCATTCCGAATCCGCCCT 840  
TCTGATCGTATGTACGTCTCGATGCCAATTTATCACACTG 880  
CAGCTGGAATTCTTGGAGTTGGGCAAGCTCTGTTGGGTGG 920  
ATCATCGTGTGTCATTAGAAAAAAATTCTCGGCTAGCAAC 960  
TTTTGGAGGGATTGTGTAAAGTATGATTGTACAGTTTCAC 1000  
1010 1020 1030 1040  
AATACATTGGAGAGATTTGTCTGGTACTTGTTGGCTCAGCC 1040  
AGTTGTGGAAGAGGAATCCAGGCATAGAATGAGATTGTTG 1080  
GTTGGAAACGGACTCCGTGCTGAAATCTGGCAACCATTG 1120  
TAGATCGATTCCGTGTCAGAATTGGAGAAGTTTATGGTTC 1160  
AACTGAAGGAACATTCATCTCTCGTGAACATTGACGGACAT 1200

FIG. 76A

FIG. 76B

10	20	30	40	
<hr/>				
MKLEELVTVM	LLTVAVIAQN	LPIGVILAG	VLILYITV	VHG 40
DFIYRSYL	TLNRDLTGL	ALIEVKIDL	WWRLHONK	GIHEL 80
FLDIVKKN	PNKPAMIDI	ETNTTET	YAEFNAHC	NRANYFO 120
GLGYRSGD	VVALYMENS	VEFVAAWM	GGLAKIGV	VTAWINSN 160
LKREQLVH	CITASKTKA	IITSVTLQ	NIMLDAID	QKLFDVE 200
210	220	230	240	
<hr/>				
GIEVYSVG	EPKKNSGF	KNLKKKL	DAQITTEP	KTLDIVDFK 240
SILCFIYT	SGTTGMPK	AAVMKHFR	YYSIAVGA	AKSFGIRP 280
SDRMYSMP	IYHTAAGI	LGVGOALL	GGSSCVIR	KKFSASN 320
FWRDCVKY	DCTVSQYI	GEICRYLL	AQPVVEES	SRHRMRL 360
VGNGLRAE	IWQPFVDR	FRVRIGEL	YGSTEGT	SSLVNIDGH 400
410	420	430	440	
<hr/>				
VGACGFLP	ISPLTKKM	HPVRLIKV	DDVTGEA	IRTSGLCI 440
ACNPGESG	AMVSTIRK	NNPLLQF	EGYLNKK	ETNKKIIRDV 480
FAKGDSCF	LTGDLHWD	RLGYVYF	KDRTGDT	FRWKGENVS 520
TTEVEAIL	HPITGLSD	ATVYGVE	VPQREGR	VGMASVVRVY 560
SHEEDETQ	FVHRVGAR	LASSLT	SYAIPQF	MRICQDVEKTG 600
610	620	630	640	
<hr/>				
TFKLVKTN	LQRLGIMD	APSOSIYI	YNSENRN	FVPFDNDLR 640
CKVSLGSY	PF			650

FIG. 77

FIG. 78A



1210 1220 1230 1240  
GTGGATAACCATGTTGGAGCTTGTGGATTTCATGCCAATTT 1240  
ATCCCCATATTGGATCCCTCTACCCAGTTCGACTTATTAA 1280  
GGTTGATAGAGCCACTGGAGAGCTTGAACGTGATAAGAAC 1320  
GGACTCTGTGTGCCGTGTGTGCCTGGTGAAACTGGGGAAA 1360  
TGGTTGGCGTTATCAAGGAGAAAGATATTCTTCTAAAGTT 1400  
1410 1420 1430 1440  
CGAAGGATATGTCAGCGAAGGGGATACTGCAAAGAAAATC 1440  
TACAGAGATGTGTTCAAGCATGGAGATAAGGTGTTTGCAA 1480  
GTGGAGATATTCTTCATTGGGATGATCTTGGATACTTGTA 1520  
CTTTGTGGACCGTTGTGGAGACACTTTCCGTTGGAAAGGG 1560  
GAGAACGTGTCAACTACTGAAGTTGAGGGAATTCTTCAGC 1600  
1610 1620 1630 1640  
CTGTGATGGATGTGGAAGATGCAACTGTTTATGGAGTCAC 1640  
TGTCGGTAAAATGGAGGGGCGTGCCGGAATGGCTGGTATT 1680  
GTCGTCAAGGATGGAACGGATGTTGAGAAATTCATCGCCG 1720  
ATATTACTTCTCGACTGACCGAAAATCTGGCGTCTTACGC 1760  
AATCCCTGTTTTTCATTTCGGCTGTGCAAGGAAGTTGATCGA 1800  
1810 1820 1830 1840  
ACCGGAACCTTCAAACCTCAAGAAGACTGATCTTCAAAAAC 1840  
AAGGTTACGACCTGGTTGCTTGTAAGGAGACCCAATTTA 1880  
CTACTGGTCAGCTGCAGAAAAATCCTACAAACCACTGACT 1920  
GACAAAATGCAACAGGATATTGACACTGGTGTGTTATGATC 1960  
GCATTTAA 1968

FIG. 78B

10	20	30	40
MREMPDSPKFALVTFVVYAVVLYNVNSVFWKFVFIGYVVF 40			
RLLRTDFGRRALATLPRDFAGLKLLISVKSTIRGLFKKDR 80			
PIHEIFLNQVKQHPNKVAIIIEIESGRQLTYQELNALANQY 120			
ANLYVSEGYKMGDVVALFMENSIDFFAIWLGLSKIGVSA 160			
FINSNLKLEPLAHSINVSKCKSCITNINLLPMFKAAREKN 200			
210	220	230	240
LISDEIHVFLAGTQVDGRHRSLOQDLHLFSEDEPPVIDGL 240			
NFRSVLCYIYTSGTTGNPKPAVIKHFYFWIAMGAGKAFG 280			
INKSDVVYITMPMYHSAAGIMGIGSLIAFGSTAVIRKKFS 320			
ASNFWKDCVKYNVTATQYIGEICRYLLAANPCPEEKQHN 360			
RLMWGNGLRGQIWKEFVGRFGIKKIGELYGSTEGNSNIVN 400			
410	420	430	440
VDNHVGACGFMPYYPHIGSLYPVRLIKVD RATGE LERDKN 440			
GLCVPCVPGETGEMVGVIKEKDILLKFEGYVSEGDTAKKI 480			
YRDVFKHGDVKVFASGDILHWDDLGYLYFVDRCGDTFRWKG 520			
ENVSTTEVEGILQPVMDVEDATVYGVTVGKMEGRAGMAGI 560			
VVKDGT DVEKFIADITSRLTENLASYAIPVFI RLCKEVD R 600			
610	620	630	640
TGTFKLKKTDLQKQGYDLVACKGDP IYYWSAAEKS YKPLT 640			
DKMQQDIDTG VYDRI 655			

FIG. 79

10 20 30 40

ATGGCGTGATGCATCAGGCTCAGCTATAAATGATCTAG 40  
AGGAATTGCTAACTGGTCCATCAGTACCCATCGTTGCTGG 80  
AGCTGCTGGAGCTGCAGCTCTCACTGCCTACATTAACGCC 120  
AAATACCACATAGCCCATGATCTCAAGACCCTCGGTGGTG 160  
GATTGACACAATCGTCCGAAGCGATTGATTTTCATAAACCG 200

210 220 230 240

CCGCGTCGCACAAAAGCGCGTCTCACGCACCACATCTTC 240  
CAGGAGCAGGTCCAAAAACAATCAAATCATCCCTTTCTTA 280  
TCTTTGAGGGCAAGACATGGTCTTACAAGGAGTTCTCTGA 320  
GGCATAACGAGGGTCGCGAACTGGCTGATTGATGAGCTG 360  
GACGTACAAGTAGGGGAGATGGTCGCAATTGATGGCGGAA 400

410 420 430 440

ATAGTGAGAGCACCTGATGCTTTGGCTTGCACTTGATGC 440  
AATCGGTGCGGCTACGAGTTTTTTGAACTGGAACCTGACA 480  
GGGGCAGGGTTAATTCATTGCATAAAGCTATGCGAATGTC 520  
GATTGTTATCGCAGACATCGATATTAAAGCGAACATTGA 560  
ACCGTGCCGTGGCGAACTGGAGGAGACGGGCATCAACATT 600

610 620 630 640

CACTACTATGACCCATCCTTCATCTCATCGCTACCGAATA 640  
ACACGCCAATTCCCGACAGCCGCACTGAGAACATTGAATT 680  
AGATTCAGTACGAGGACTGATAACACATCTGGAACCACT 720  
GGTCTACCTAAAGGCGTGTTTATAAGCACTGGCCGCGAGC 760  
TTAGGACTGACTGGTCGATTTCAAAGTATCTAAATCTCAA 800

810 820 830 840

GCCACGGATCGAATGTATACATGTATGCCGCTCTACCAT 840  
GCCGCTGCACACAGCCTCTGTACAGCATCAGTTATTCATG 880  
GTGGAGGTACCGTGGTATTGAGCAGGAAATTTCTCACACAA 920  
GAAGTTCTGGCCTGAAGTTGTGGCTTCGGAAGCAAATATC 960  
ATTCAGTACGTTGGTGAATTAGGTCGATATCTCCTGAATG 1000

1010 1020 1030 1040

GTCCAAAGAGTCCTTACGACAGGGCCCATAAAGTCCAGAT 1040  
GGCGTGGGGCAATGGCATGCGTCCAGACGTGTGGGAAGCG 1080  
TTTCGTGAACGCTTCAACATACCAATTATTCATGAGCTCT 1120  
ATGCCGCAACCGATGGGCTCGGGTCAATGACCAATCGTAA 1160  
CGCGGGGCCCTTTTACAGCAAACCTGTATTGCGCTGCGAGGG 1200

FIG. 80A

1210 1220 1230 1240  
CTGATCTGGCACTGGAAATTTTCGAAATCAGGAAGTGCTGG 1240  
TCAAGATGGATCTCGATACTGATGAGATCATGAGAGATCG 1280  
CAATGGGTTTGGGATACGATGCGCTGTCAATGAACCTGGA 1320  
CAGATGCTTTTTTCGGCTGACACCCGAAACTCTGGCTGGTG 1360  
CACCAAGCTACTACAACAACGAAACGGCCACACAGAGCAG 1400  
1410 1420 1430 1440  
GCGGATTACAGATGTGTTTCAAAGGGTGACCTGTGGTTC 1440  
AAGTCCGGTGACATGCTACGGCAAGACGCCGAAGGCCGCG 1480  
TCTACTTTGTCGATCGACTAGGCGATACGTTCCGCTGGAA 1520  
ATCCGAAAACGTTTCTACCAATGAAGTCGCGGACGTGATG 1560  
GGCACATTTCTCAGATTGCTGAAACGAATGTATACGGTG 1600  
1610 1620 1630 1640  
TCCTTGTGCCGGGTAACGATGGTTCGAGTGCGCAGCCTCAA 1640  
TTGTCATGGCAGACGGCGTGACAGAGTCGACATTCGCTTC 1680  
GCTGCCCTTGCAAAGCACGCCCGAGATCGGTTACCGGGTT 1720  
ATGCTGTACCACTGTTTCTGAGGGTAACTCCAGCACTTGA 1760  
ATATACGGGCACATTAAAGATTTCAGAAAGGACGCCTCAAG 1800  
1810 1820 1830 1840  
CAGGAAGGTATAGACCCAGATAAGATTTCCGGCGAAGATA 1840  
AGTTATACTGGCTGCCGCCTGGTAGCGATATATATTTACC 1880  
ATTTGGAAAGATGGAGTGGCAGGGAATTGTAGATAAGCGT 1920  
ATACGGCTGTGA 1932

FIG. 80B

10	20	30	40
MACMHQAQLYNDLEELLTGPSVPIVAGAAGAAALTAYINA 40			
KYHIAHDLKTLGGGLTQSSAIDFINRRVAQKRVLTHHIF 80			
QEQVQKQSNHPFLIFEGKTWSYKEFSEAYTRVANWLIDEL 120			
DVQVGEMVAIDGGNSAEHLMLWLALDAIGAATSFLNWNLT 160			
GAGLIHCICKLCECRFVIADIDIKANIEPCRGELEETGINI 200			
210	220	230	240
HYYDPSFISSLPNNTPIPDSRTENIELDSVRGLIYTS GTT 240			
GLPKGVFI STGRELRTDWSISKYLNKPTDRMYTCMP LYH 280			
AAAHSLCTASVIHGGGT VVLSRKFSHKKFWPEV VASEANI 320			
IQYVGELGRYLLNGPKSPYDRAHKVQMAWGNGMRPDV WEA 360			
FRERFNIPIIHEL YAATDGLGSM TNRNAGPFTANCIALRG 400			
410	420	430	440
LIWHWKFRNQEV LVKMDLDTDEIMRDRNGFAIRCAV NEPG 440			
QMLFRLTPETLAGAPSYNNETATQSRRI TDVFQKGDLWF 480			
KSGDMLRQDAEGRVYFVDRLGDTFRWKSENVSTNEV ADVM 520			
GTFPQIAETNVYGVLPVPGNDGRVRS LNCHGRRRDRVD IRF 560			
AALAKHARDRLPGYAVPLFLRVTPALEYTGTLKIQKGR LK 600			
610	620	630	640
QEGIDPKISGEDKLYWLPPGSDIYLPFGKMEWQGI VDKR 640			
IRL 643			

FIG. 81

10 20 30 40  
 CTTTACCATTTCATCAGCTTCATTCTGCATTTTTAGCTTGA 40  
 CGGCAGCCGGGTCTACGCTGATCATCGGCCGCAAGTTCTC 80  
 CGCGAGAAACTTCATAAAGGAAGCGCGGAGAACGACGCC 120  
 ACGGTCATCCAGTACGTGGGTGAGACCTTGCGATATCTGC 160  
 TCGCCACCCCCGGTGAAACCGATCCAGTTACTGGCGAAGA 200  
 210 220 230 240  
 CCTGGACAAAAAGCACAAATATTCGAGCAGTATACGGCAAC 240  
 GGGCTACGGCCGGATATCTGGAACCGCTTCAAGGAGCGCT 280  
 TCAACGTGCCGACGGTTGCCGAATTTTATGCTGCAACCGA 320  
 GAGCCCAGGCGGAACATGGAACCTATTCAACAAATGACTTC 360  
 ACTGCCGGAGCCATTGGGCACACTGGCGTGCTTAGTGGAT 400  
 410 420 430 440  
 GGCTTCTTGGACGCGGCCTTACTATTGTGCGAGGTGGACCA 440  
 GGAATCACAGGAACCATGGCGCGATCCCCAAACCGGGTTC 480  
 TGCAAGCCGGTCCCGCGAGGCGAAGCAGGCGAGCTCCTGT 520  
 ATGCCATTGATCCGGCCGACCCGGGCGAGACCTTCCAGGG 560  
 CTA CTACCGCAACTCCTTTAGAGCACACTGGCGGCCG 597

FIG. 82

10 20 30 40  
 LYHSSASFCIFSLTAAGSTLIIGRKFSARNFIKEARENDA 40  
 TVIQYVGETLRYLLATPGETDPVTGEDLDKKNIRAVYGN 80  
 GLRPDIWNRFKERFNVPTVAEFYAATESPGGTWNYSTNDF 120  
 TAGAIGHTGVLSGWLLGRGLTIVEYDQESQEPWRDPOTGF 160  
 CKPVPRGEAGELLYAIDPADPGETFQGYRNSFRAHWRP 199

FIG. 83

10 20 30 40  
 GCAAAGGCCGACGCGTGCGCTGCGGACGGGTAACGTGATCA 40  
 GGGCGGACAACGAAGGGCGACTCTTCTTCCACGACCGGAT 80  
 CGGAGACACGTTCCGATGGAAGGGAGAGACNGTCAGCACA 120  
 CAAGAGGTCAGTTTGGTGCTCGGACGACACGACTCAATCA 160  
 AGGAGGCCAACGTGTACGGCGTGACGGTGCCGAACCACGA 200  
 210 220 230 240  
 CGGGCGGGCCGGCTGCGCTGCGCTCACGCTATCAGACGCT 240  
 CTGGCGACTGAAAAGAAGCTGGGCGATGAGCTGCTAAAGG 280  
 GATTGGCTACTCACTCGTTCGACTTCGCTTCCCAAGTTTGC 320  
 GGTGCCGCAGTTCCTACGGGTGGTGCGCGGCGAGATGCAG 360  
 TCAACGGGCACCAACAAGCAACAGAAGCACGACCTGAGGG 400  
 410 420 430 440  
 TGCAGGGTGTAGAGCCGGGCAAGGTGGGCGTAGACGAGGT 440  
 GTACTGGTTGCGGGGAGGGACATATGTACCATTTCGGAACA 480  
 GAGGATTGGGATGGGTTGAAGAAGGGTCTTGTGAAGTTGT 520  
 GA 522

FIG. 84

10 20 30 40  
 AKADAWLRTGNVIRADNEGRLFFHDIRIGDTFRWKGETVST 40  
 QEVSLVLGRHDSIKEANYVGTVPNHDGRAGCAALTLSDA 80  
 LATEKKLGDELLKGLATHSSTSLPKFAVPOFLRVVRGEMQ 120  
 STGTNKQOKHDLRVQGVPEPGKVGVDVYWLRGGETYVPFGT 160  
 EDWDGLKKGLVKL 173

FIG. 85

10	20	30	40
-----			
ATGTCTCCCATACAGGTTGTTGTCTTTGCCTTGTCAAGGA	40		
TTTTCTTGCTATTATTCAGACTTATCAAGCTAATTATAAC	80		
CCCTATCCAGAAATCACTGGGTTATCTATTTGGTAATTAT	120		
TTTGATGAATTAGACCGTAAATATAGATACAAGGAGGATT	160		
GGTATATTATTCCCTTACTTTTGAAGGCGTGTGTTTGTTA	200		
210	220	230	240
-----			
TATCATTGATGTGAGAAGACATAGGTTTCAAACTGGTAC	240		
TTATTTATTAACAGGTCCAACAAAATGGTGACCATTTAG	280		
CGATTAGTTACACCCGTCCCATGGCCGAAAAGGGAGAATT	320		
TCAACTCGAAACCTTTACGTATATTGAACTTATAACATA	360		
GTGTTGAGATTGTCTCATATTTTGCATTTTGATTATAACG	400		
410	420	430	440
-----			
TTCAGGCCGGTGACTACGTGGCAATCGATTGTACTAATAA	440		
ACCTCTTTTCGTATTTTTATGGCTTTCTTTGTGGAACATT	480		
GGGGCTATTCCAGCTTTTTTAACTATAATACTAAAGGCA	520		
CTCCGCTGGTTCACTCCCTAAAGATTTCCAATATTACGCA	560		
GGTATTTATTGACCCTGATGCCAGTAATCCGATCAGAGAA	600		
610	620	630	640
-----			
TCGGAAGAAGAAATCAAAAACGCACTTCCTGATGTTAAAT	640		
TAAACTATCTTGAAGAACAAGACTTAATGCATGAACCTTTT	680		
AAATTCGCAATCACCGGAATTCTTACAACAAGACAACGTT	720		
AGGACACCACTAGGCTTGACCGATTTTAAACCCTCTATGT	760		
TAATTTATACATCTGGAACCACTGGTTTGCCTAAATCCGC	800		
810	820	830	840
-----			
TATTATGTCTTGGAGAAAATCCTCCGTAGGTTGTCAAGTT	840		
TTTGGTCATGTTTTACATATGACTAATGAAAGCACTGTGT	880		
TCACAGCCATGCCATTGTTCCATTCAACTGCTGCCTTATT	920		
AGGTGCGTGCGCCATTCTATCTCACGGTGGTTGCCTTGCG	960		
TTATCGCATAAATTTTCTGCCAGTACATTTTGGAAGCAAG	1000		
1010	1020	1030	1040
-----			
TTTATTTAACAGGAGCCACGCACATCCAATATGTCCGAGA	1040		
AGTCTGTAGATACCTGTTACATACGCCAATTTCTAAGTAT	1080		
GAAAAGATGCATAAGGTGAAGGTTGCTTATGGTAACGGGC	1120		
TGAGACCTGACATCTGGCAGGACTTCAGGAAGAGGTTCAA	1160		
CATAGAAGTTATTGGTGAATTCTATGCCGCAACTGAAGCT	1200		

FIG. 86A



1210 1220 1230 1240  
CCTTTTGCTACAACCTACCTTCCAGAAAGGTGACTTTGGAA 1240  
TTGGCGCATGTAGGAACTATGGTACTATAATTCAATGGTT 1280  
TTTGTCATTCCAACAAACATTGGTAAGGATGGACCCAAAT 1320  
GACGATTCCGTTATATATAGAAATTCCAAGGGTTTCTGCG 1360  
AAGTGGCCCCTGTTGGCGAACCAGGAGAAATGTTAATGAG 1400  
1410 1420 1430 1440  
AATCTTTTTCCCTAAAAAACCAGAAACATCTTTTCAAGGT 1440  
TATCTTGGTAATGCCAAGGAAACAAAGTCCAAAGTTGTGA 1480  
GGGATGTCTTCAGACGTGGCGATGCTTGGTATAGATGTGG 1520  
AGATTTATTAAAAGCGGACGAATATGGATTATGGTATTTT 1560  
CTTGATAGAATGGGTGATACTTTCAGATGGAAATCTGAAA 1600  
1610 1620 1630 1640  
ATGTTTCCACTACTGAAGTAGAAGATCAGTTGACGGCCAG 1640  
TAACAAAGAACAATATGCACAAGTTCTAGTTGTTGGTATT 1680  
AAAGTACCTAAATATGAAGGTAGAGCTGGTTTTGCAGTTA 1720  
TTAAACTAACTGACAACTCTCTTGACATCACTGCAAAGAC 1760  
CAAATTATTAAATGATTCCTTGAGCCGGTTAAATCTACCG 1800  
1810 1820 1830 1840  
TCTTATGCTATGCCCCTATTTGTTAAATTTGTTGATGAAA 1840  
TTAAATGACAGATAACCTCATAAAATTTTGA 1872

FIG. 86B

FIG. 87

10	20	30	40
----- ----- ----- -----			
GTGTCCGATTACTACGGCGGGCGCACACACAACGGTCAGGC	40		
TGATCGACCTGGCAACTCGGATGCCGCGAGTGTTGGCGGA	80		
CACGCCGGTGATTGTGCGTGGGGCAATGACCGGGCTGCTG	120		
GCCCGGCCGAATTCCAAGGCGTCGATCGGCACGGTGTTCC	160		
AGGACCGGGCGCTCGCTACGGTGACCGAGTCTTCCTGAA	200		
210	220	230	240
----- ----- ----- -----			
ATTCGGCGATCAGCAGCTGACCTACCGCGACGCTAACGCC	240		
ACCGCCAACCGGTACGCCGCGGTGTTGGCCGCCCGCGGCG	280		
TCGGCCCCGGCGACGTCGTTGGCATCATGTTGCGTAACTC	320		
ACCCAGCACAGTCTTGGCGATGCTGGCCACGGTCAAGTGC	360		
GGCGCTATCGCCGGCATGCTCAACTACCACCAGCGCGGCG	400		
410	420	430	440
----- ----- ----- -----			
AGGTGTTGGCGCACAGCCTGGGTCTGCTGGACGCGAAGGT	440		
ACTGATCGCAGAGTCCGACTTGGTCAGCGCCGTCGCCGAA	480		
TGCGGCGCCTCGCGCGGCCGGGTAGCGGGCGACGTGCTGA	520		
CCGTGAGGACGTGGAGCGATTGCCACAACGGCGCCCGC	560		
CACCAACCCGGCGTCGGCGTCGGCGGTGCAAGCCAAAGAC	600		
610	620	630	640
----- ----- ----- -----			
ACCGCGTTCTACATCTTCACCTCGGGCACCACCGGATTTTC	640		
CCAAGGCCAGTGTCATGACGCATCATCGGTGGCTGCGGGC	680		
GCTGGCCGTCTTCGGAGGGATGGGGCTGCGGCTGAAGGGT	720		
TCCGACACGCTCTACAGCTGCCTGCCGCTGTACCACAACA	760		
ACGCGTTAACGGTCGCGGTGTGCTCGGTGATCAATTCTGG	800		
810	820	830	840
----- ----- ----- -----			
GGCGACCCTGGCGCTGGGTAAGTCGTTTTTCGGCGTCGCGG	840		
TTCTGGGATGAGGTGATTGCCAACCGGGCGACGGCGTTTCG	880		
TCTACATCGGCGAAATCTGCCGTTATCTGCTCAACCAGCC	920		
GGCCAAGCCGACCGACCGTCCCCACCAGGTGCGGGTGATC	960		
TGCGGTAACGGGCTGCGGCCGGAGATCTGGGATGAGTTCA	1000		
1010	1020	1030	1040
----- ----- ----- -----			
CCACCCGCTTCGGGGTCGCGCGGGTGTGCGAGTTCTACGC	1040		
CGCCAGCGAAGGCAACTCGGCCTTTATCAACATCTTCAAC	1080		
GTGCCCAGGACCGCCGGGGTATCGCCGATGCCGCTTGCCCT	1120		
TTGTGGAATACGACCTGGACACCGGCGATCCGCTGCGGGA	1160		
TGCGAGCGGGCGAGTGCGTCGGGTACCCGACGGTGAACCC	1200		

FIG. 88A

1210 1220 123  
 GGCCTGTTGCTTAGCCGGGTCAACCGGCTGCAGCCGTTTCG 1240  
 ACGGCTACACCGACCCGGTTGCCAGCGAAAAGAAGTTGGT 1280  
 GCGCAACGCTTTTCGAGATGGCGACTGTTGGTTCAACACC 1320  
 GGTGACGTGATGAGCCCGCAGGGCATGGGCCATGCCGCCT 1360  
 TCGTCGATCGGCTGGGCGACACCTTCCGCTGGAAGGGCGA 1400  
 1410 1420 1430 1440  
 GAATGTCGCCACCACTCAGGTCGAAGCGGCACTGGCCTCC 1440  
 GACCAGACCGTTCGAGGAGTGCACGGTCTACGGCGTCCAGA 1480  
 TTCCGCGCACCCGGCGGGCGCGCCGGAATGGCCGCGATCAC 1520  
 ACTGCGCGCTGGCGCCGAATTCGACGGCCAGGCGCTGGCC 1560  
 CGAACGGTTTACGGTCACTTGCCCGGCTATGCACTTCCGC 1600  
 1610 1620 1630 1640  
 TCTTTGTTCTGGGTAGTGGGGTCGCTGGCGCACACCACGAC 1640  
 GTTCAAGAGTCGCAAGGTGGAGTTGCGCAACCAGGCCTAT 1680  
 GGCGCCGACATCGAGGATCCGCTGTACGTACTGGCCGGCC 1720  
 CGGACGAAGGATATGTGCCGTACTACGCCGAATACCCTGA 1760  
 GGAGGTTTCGCTCGGAAGGCGACCGCAGGGCTAG 1794

FIG. 88B

10 20 30 40  
 MSDYYGGAHTTVRLIDLATRMPRVLADTPVIVRGAMTGLL 40  
 ARPNSKASIGTVFQDRAARYGDRVFLKFGDQQLTYRDANA 80  
 TANRYAAVLAARGVGPGDVVGIMLRNSPSTVLA MLATVKC 120  
 GAIAGMLNYHQGEVLHSLGLLDAKVLI AESDLVSAVAE 160  
 CGASRGRVAGDVLTVEDVERFATTAPATNPASASAVQAKD 200  
 210 220 230 240  
 TAFYIFTSGTTGFPKASVMTHHRWLRALAVFGGMGLRLKG 240  
 SDTLYSCLPLYHNNALTVAVSSVINSGATLALGKSFSASR 280  
 FWDEVIANRATAFVYIGEICRYLLNQPAKPTDRAHQVRVI 320  
 CGNGLRPEIWDEFTTRFGVARVCEFYAASEGNSAFINIFN 360  
 VPRTAGVSPMPLAFVEYDLDTGDLPLRDASGRVRRVPDGEP 400  
 410 420 430 440  
 GLLLSRVNRLQPFDDGYTDPVASEKKLVRNAFRDGDGWFNT 440  
 GDVMSPQGMGHA AFVDRLGDTFRWKGENVATTQVEAALAS 480  
 DQTVEECTVYGVQIPRTGGRAGMAAITLRAGAEFDGQALA 520  
 RTVYGHLPGYALPLFVRVVGSLAHTTTFKSRKVELRNQAY 560  
 GADIEDPLYVLAGPDEGYVPYYAEYPEEVSLGRRPQG 597

FIG. 89

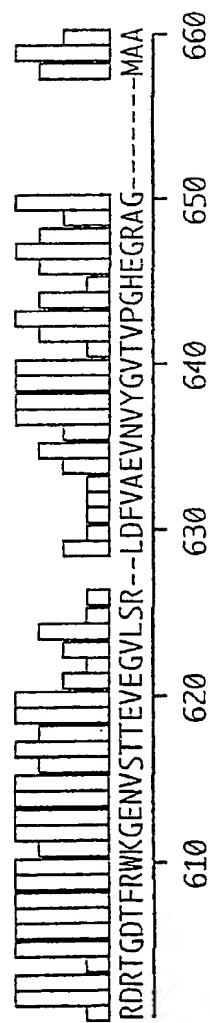
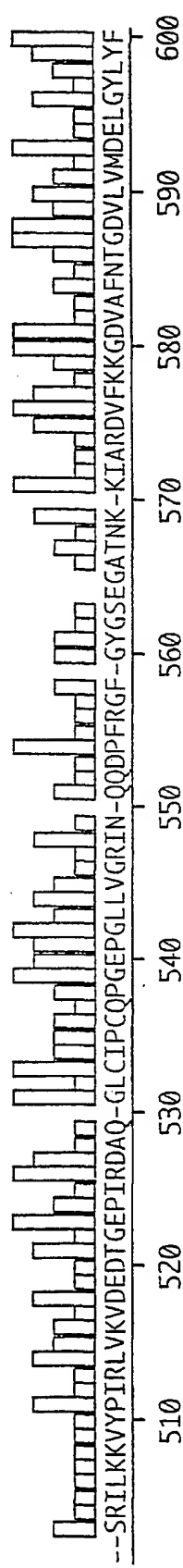
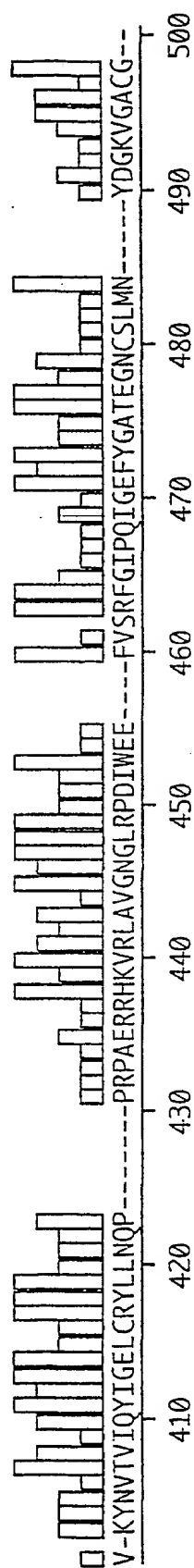
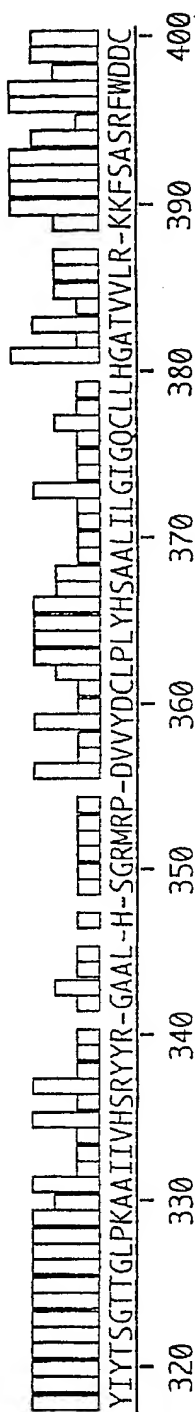


FIG. 90

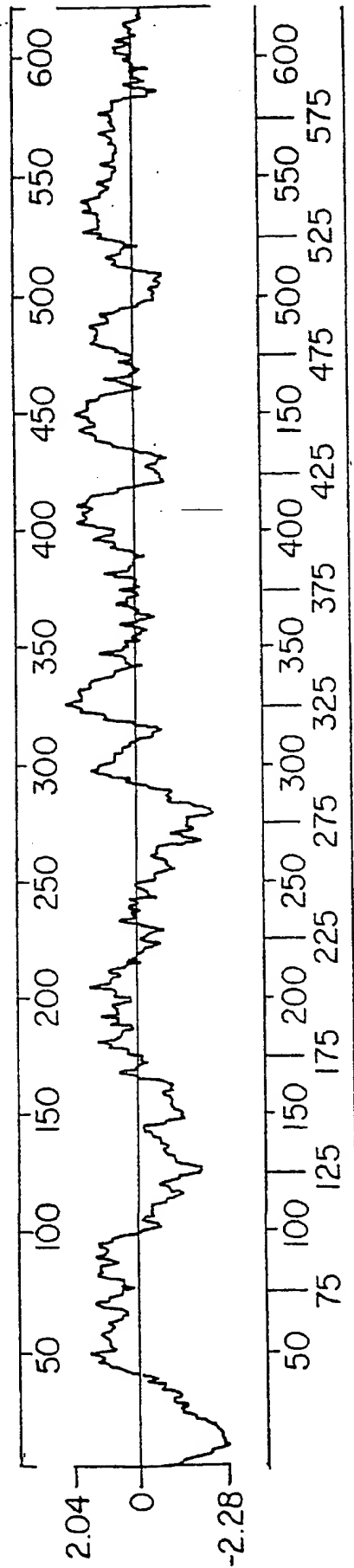


FIG. 91

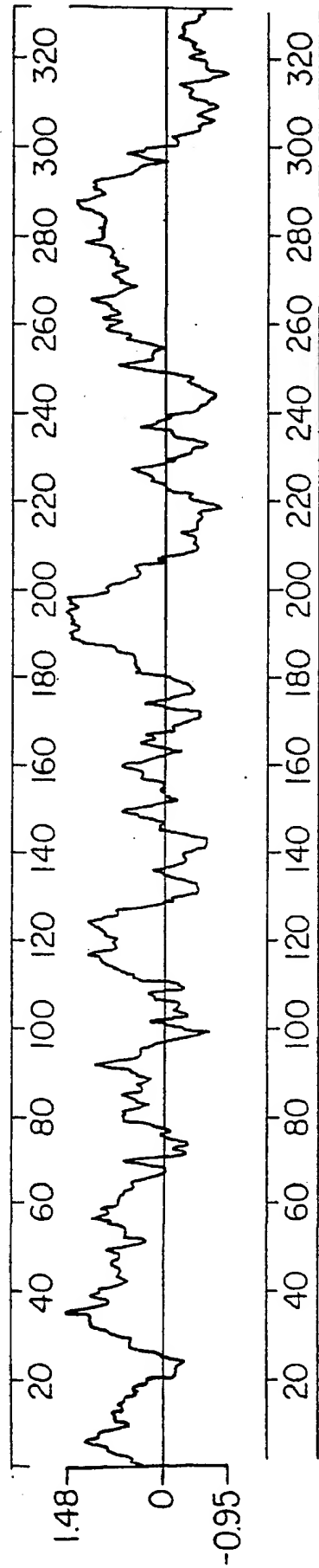


Figure 92

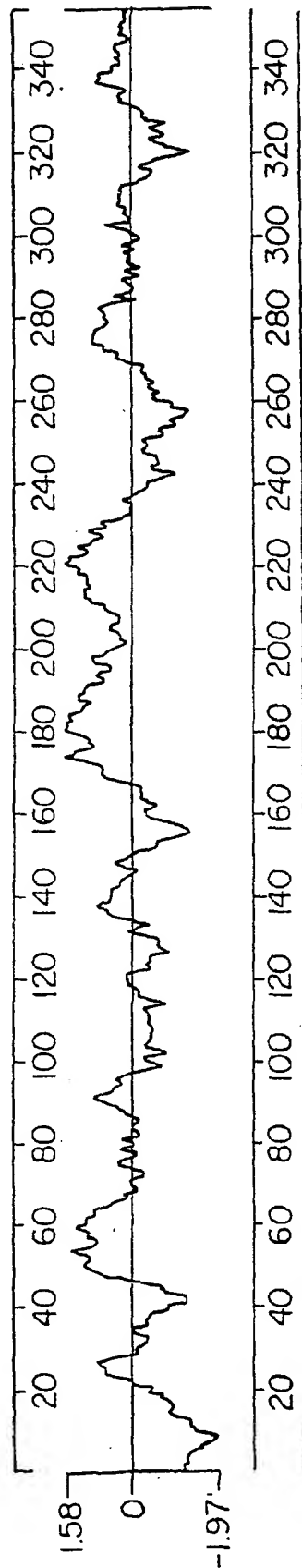


Figure 93

1	cga ccc acg cgt ccg ggg atg ttt gcg agc
1	M F A S
31	ggc tgg aac cag acg gtg ccg ata gag gaa
5	G W N Q T V P I E E
61	gcg ggc tcc atg gct gcc ctc ctg ctg ctg
15	A G S M A A L L L L
91	ccc ctg ctg ctg ttg cta ccg ctg ctg ctg
25	P L L L L L P L L L
121	ctg ctg aag cta cac ctc tgg ccg cag ttg
35	L L K L H L W P Q L
151	cgc tgg ctt ccg gcg gac ttg gcc ttt gcg
45	R W L P A D L A F A
181	gtg cga gct ctg tgc tgc aaa agg gct ctt
55	V R A L C C K R A L
211	cga gct cgc gcc ctg gcc gcg gct gcc gcc
65	R A R A L A A A A A
241	gac ccg gaa ggt ccc gag ggg ggc tgc agc
75	D P E G P E G G C S

FIG. 94A



271     ctg gcc tgg cgc ctc gcg gaa ctg gcc cag  
85       L   A   W   R   L   A   E   L   A   Q

301     cag cgc gcc gcg cac acc ttt ctc att cac  
95       Q   R   A   A   H   T   F   L   I   H

331     ggc tgc cgg cgc ttt agc tac tca gag gcg  
105      G   S   R   R   F   S   Y   S   E   A

361     gag cgc gag agt aac agg gct gca cgc gcc  
115      E   R   E   S   N   R   A   A   R   A

391     ttc cta cgt gcg cta ggc tgg gac tgg gga  
125      F   L   R   A   L   G   W   D   W   G

FIG. 94B

421     ccc gac ggc ggc gac agc ggc gag ggg agc  
135      P   D   G   G   D   S   G   E   G   S

451     gct gga gaa ggc gag cgg gca gcg ccg gga  
145      A   G   E   G   E   R   A   A   P   G

481     gcc gga gat gca gcg gcc gga agc ggc gcg  
155      A   G   D   A   A   A   G   S   G   A

521     gag ttt gcc gga ggg gac ggt gcc gcc aga  
165      E   F   A   G   G   D   G   A   A   R

541     ggt gga gga gag ccc gcc gcc cct ctg tca  
175      G   G   G   E   P   A   A   P   L   S

571     cct gga gca act gtg gcg ctg ctc ctc ccc  
185      P   G   A   T   V   A   L   L   L   P

601     gct ggc cca gag ttt ctg tgg ctc tgg ttc  
195      A   G   P   E   F   L   W   L   W   F

FIG. 94C

631	ggg	ctg	gcc	aag	gcc	ggc	ctg	cgc	act	gcc
205	G	L	A	K	A	G	L	R	T	A
661	ttt	gtg	ccc	acc	gcc	ctg	cgc	cgg	ggc	ccc
215	F	V	P	T	A	L	R	R	G	P
691	ctg	ctg	cac	tgc	ctc	cgc	agc	tgc	ggc	gcg
225	L	L	H	C	<u>L</u>	R	S	C	G	A
721	cgc	gcg	ctg	gtg	ctg	gcg	cca	gag	ttt	ctg
235	R	A	L	V	L	A	P	E	F	L
751	gag	tcc	ctg	gag	ccg	gac	ctg	ccc	gcc	ctg
245	E	S	L	E	P	D	L	P	A	L
781	aga	gcc	atg	ggg	ctc	cac	ctg	tgg	gct	gca
255	R	A	M	G	L	H	L	W	A	A
811	ggc	cca	gga	acc	cac	cct	gct	gga	att	agc
265	G	P	G	T	H	P	A	G	I	S
841	gat	ttg	ctg	gct	gaa	gtg	tcc	gct	gaa	gtg
275	D	L	L	A	E	V	S	A	E	V

FIG. 94D

871	gat	ggg	cca	gtg	cca	gga	tac	ctc	tct	tcc
285	D	G	P	V	P	G	Y	L	S	S
901	ccc	cag	agc	ata	aca	gac	acg	tgc	ctg	tac
295	P	Q	S	I	T	D	T	C	L	Y
931	atc	ttc	acc	tct	ggc	acc	acg	ggc	ctc	ccc
305	I	F	T	S	G	T	T	G	L	P
961	aag	gct	gct	cgg	atc	agt	cat	ctg	aag	atc
315	K	A	A	R	I	S	H	L	K	I
991	ctg	caa	tgc	cag	ggc	ttc	tat	cag	ctg	tgt
325	L	Q	C	Q	G	F	Y	Q	L	C
1021	ggt	gtc	cac	cag	gaa	gat	gtg	atc	tac	ctc
335	G	V	H	Q	E	D	V	I	Y	L

FIG. 94E

1051	gcc	ctc	cca	ctc	tac	cac	atg	tcc	ggt	tcc
345	A	L	P	L	Y	H	M	S	G	S
1081	ctg	ctg	ggc	atc	gtg	ggc	tgc	atg	ggc	att
355	L	L	G	I	V	G	C	M	G	I
1111	ggg	gcc	aca	gtg	gtg	ctg	aaa	tcc	aag	ttc
365	G	A	T	V	V	L	K	S	K	F
1141	tcg	gct	ggt	cag	ttc	tgg	gaa	gat	tgc	cag
375	S	A	G	Q	F	W	E	D	C	Q
1171	cag	cac	agg	gtg	acg	gtg	ttc	cag	tac	att
385	Q	H	R	V	T	V	F	Q	Y	I
1201	ggg	gag	ctg	tgc	cga	tac	ctt	gtc	aac	cag
395	G	E	L	C	R	Y	L	V	N	Q
1231	ccc	ccg	agc	aag	gca	gaa	cgt	ggc	cat	aag
405	P	P	S	K	A	E	R	G	H	K
1261	gtc	cgg	ctg	gca	gtg	ggc	agc	ggg	ctg	cgc
415	V	R	L	A	V	G	S	G	L	R

FIG. 94F

1291	cca gat acc tgg gag cgt ttt gtg cgg cgc
425	P D T W E R F V R R
1321	ttc ggg ccc ctg cag gtg ctg gag aca tat
435	F G P L Q V L E T Y
1351	gga ctg aca gag ggc aac gtg gcc acc atc
445	G L T E G N V A T I
1381	aac tac aca gga cag cgg ggc gct gtg ggg
455	N Y T G Q R G A V G
1411	cgt gct tcc tgg ctt tac aag cat atc ttc
465	R A S W L Y K H I F
1441	ccc ttc tcc ttg att cgc tat gat gtc acc
475	P F S L I R Y D V T
1471	aca gga gag cca att cgg gac ccc cag ggg
485	T G E P I R D P Q G
1501	cac tgt atg gcc aca tct cca ggt gag cca
495	H C M A T S P G E P
1531	ggg ctg ctg gtg gcc ccg gta agc cag cag
505	G L L V A P V S Q Q

FIG. 94G

1561	tcc cca ttc ctg ggc tat gct ggc ggg cca
515	S P F L G Y A G G P
1591	gag ctg gcc cag ggg aag ttg cta aag gat
525	E L A Q G K L L K D
1621	gtc ttc cgg cct ggg gat gtt ttc ttc aac
535	V F R P G D V F F N
1651	act ggg gac ctg <del>ctg</del> gtc tgc gat gac caa
545	T G D L L V C D D Q
1681	ggt ttt ctc cgc ttc cat gat cgt act gga
555	G F L R F H D R T G
1711	gac acc ttc agg tgg aag ggg gag aat gtg
565	D T F R W K G E N V
1741	gcc aca acc gag gtg gca gag gtc ttc gag
575	A T T E V A E V F E

FIG. 94H

1741	gcc	aca	acc	gag	gtg	gca	gag	gtc	ttc	gag
575	A	T	T	E	V	A	E	V	F	E
1771	gcc	cta	gat	ttt	ctt	cag	gag	gtg	aac	gtc
585	A	L	D	F	L	Q	E	V	N	V
1801	tat	gga	gtc	act	gtg	cca	ggg	cat	gaa	ggc
595	Y	G	V	T	V	P	G	H	E	G
1831	agg	gct	gga	atg	gca	gcc	cta	gtt	ctg	cgt
605	R	A	G	M	A	A	L	V	L	R
1861	ccc	ccc	cac	gct	ttg	gac	ctt	atg	cag	ctc
615	P	P	H	A	L	D	L	M	Q	L
1891	tac	acc	cac	gtg	tct	gag	aac	ttg	cca	cct
625	Y	T	H	V	S	E	N	L	P	P
1921	tat	gcc	cgg	ccc	cga	ttc	ctc	agg	ctc	cag
635	Y	A	R	P	R	F	L	R	L	Q

FIG. 94I

1951	gag	tct	ttg	gcc	acc	aca	gag	acc	ttc	aaa
645	E	S	L	A	T	T	E	T	F	K
1981	cag	cag	aaa	gtt	cgg	atg	gca	aat	gag	ggc
655	Q	Q	K	V	R	M	A	N	E	G
2011	ttc	gac	ccc	agc	acc	ctg	tct	gac	cca	ctg
665	F	D	P	S	T	L	S	D	P	L
2041	tac	gtt	ctg	gac	cag	gct	gta	ggc	gcc	tac
675	Y	V	L	D	Q	A	V	G	A	Y
2071	ctg	ccc	ctc	aca	act	gcc	cgg	tac	agc	gcc
685	L	P	L	T	T	A	R	Y	S	A
2101	ctc	ctg	gca	gga	aac	ctt	cga	atc	tga	gaa
695	L	L	A	G	N	L	R	I	*	
2131	ctt	cca	cac	ctg	agg	cac	ctg	aga	gag	gaa
2161	ctc	tgt								

FIG. 94J



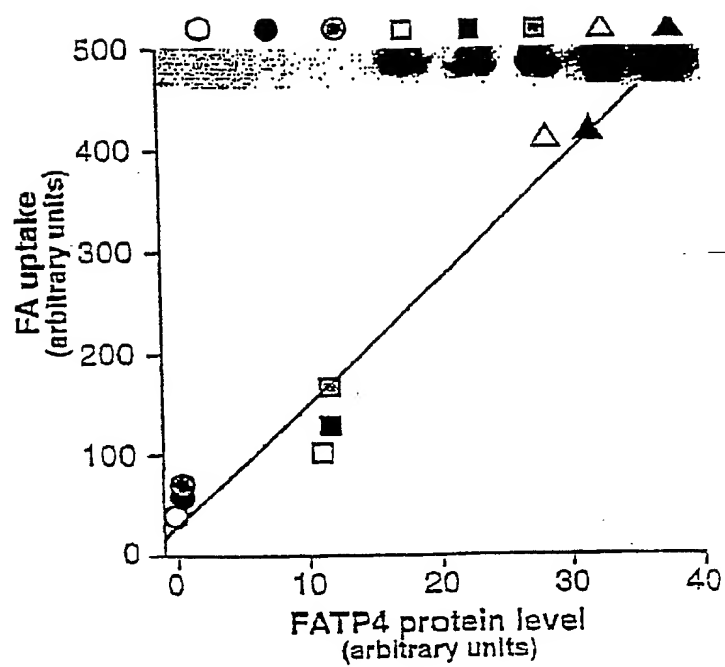


FIG. 95

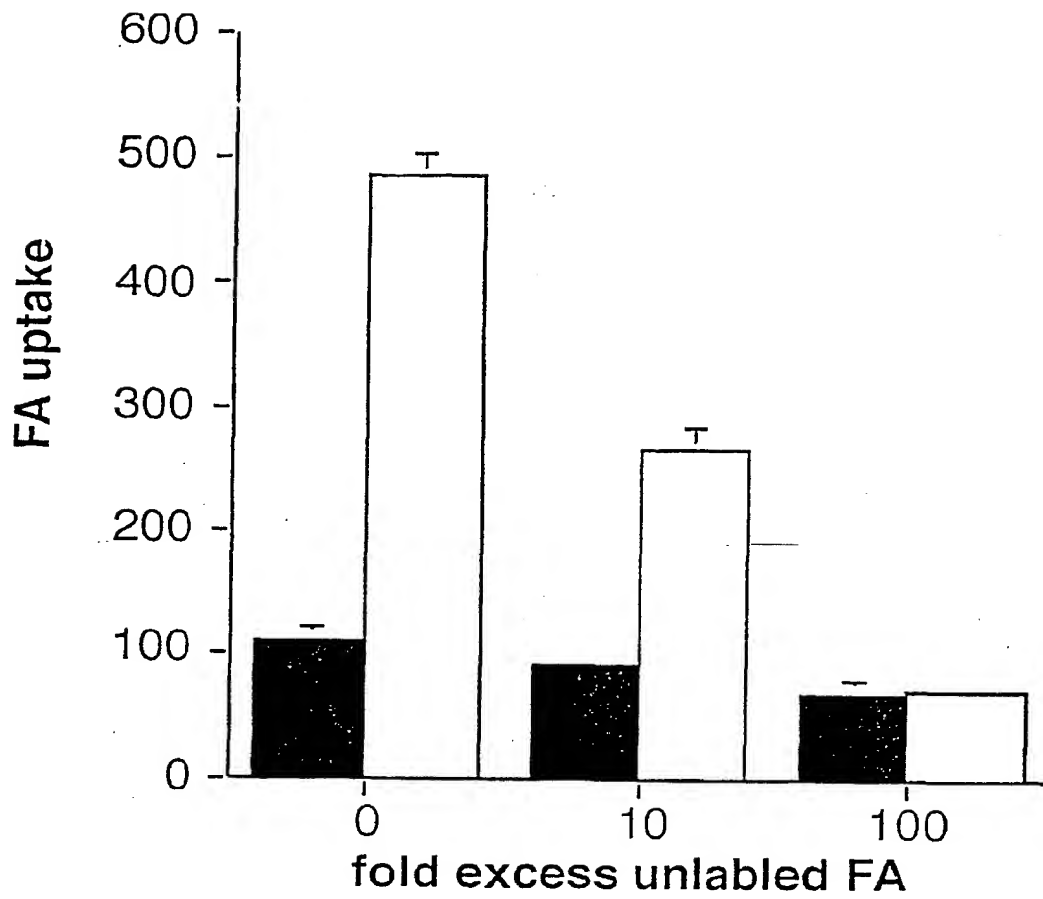


FIG. 96

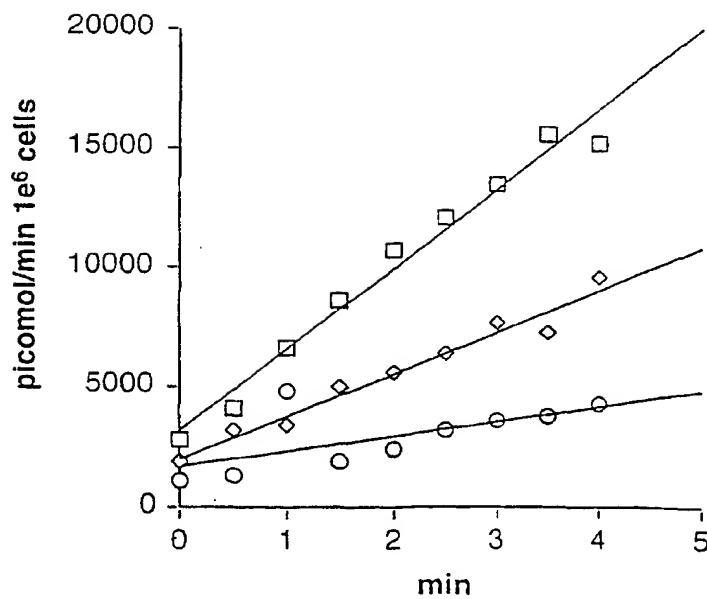


FIG. 97

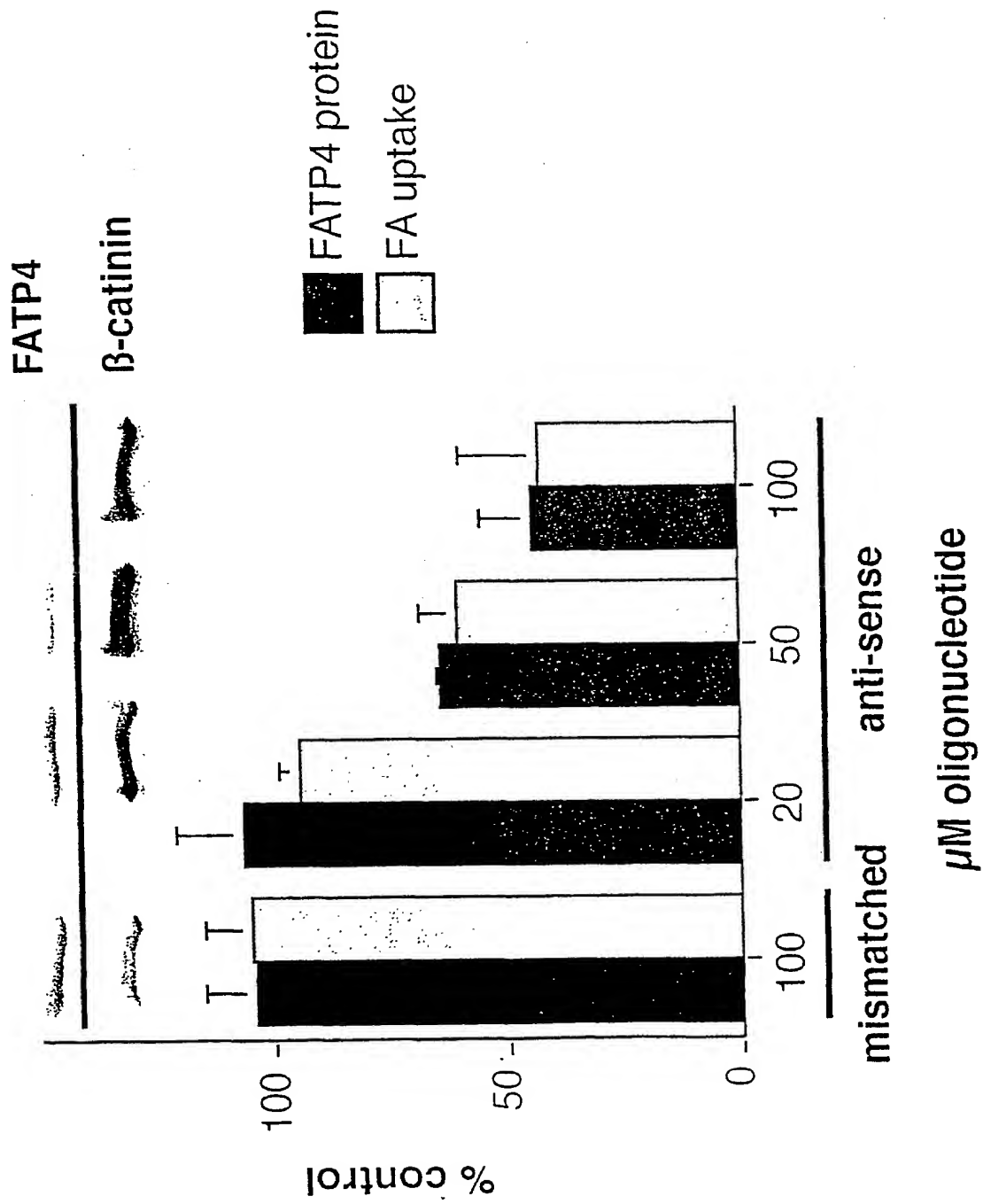


FIG. 98

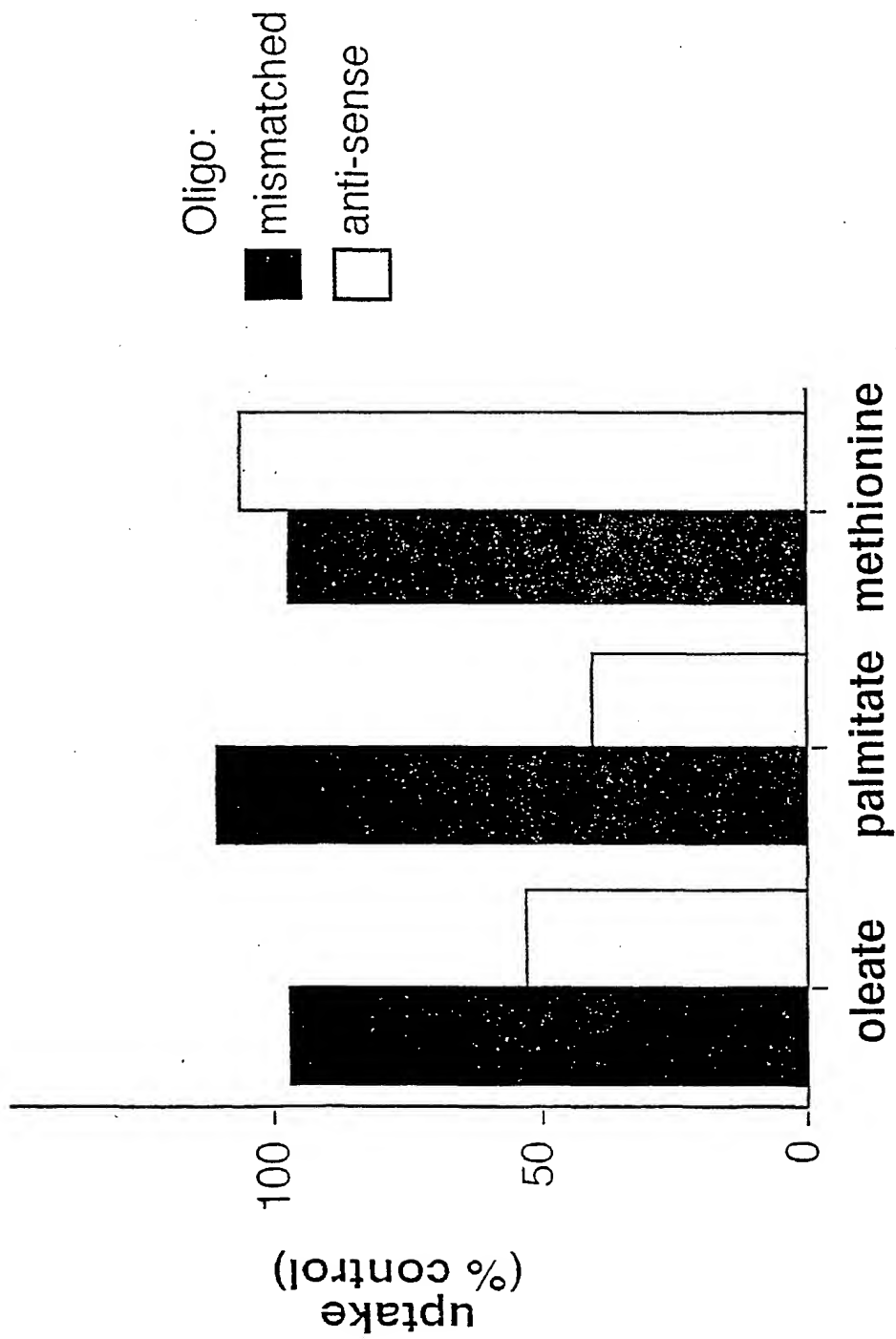


FIG. 99

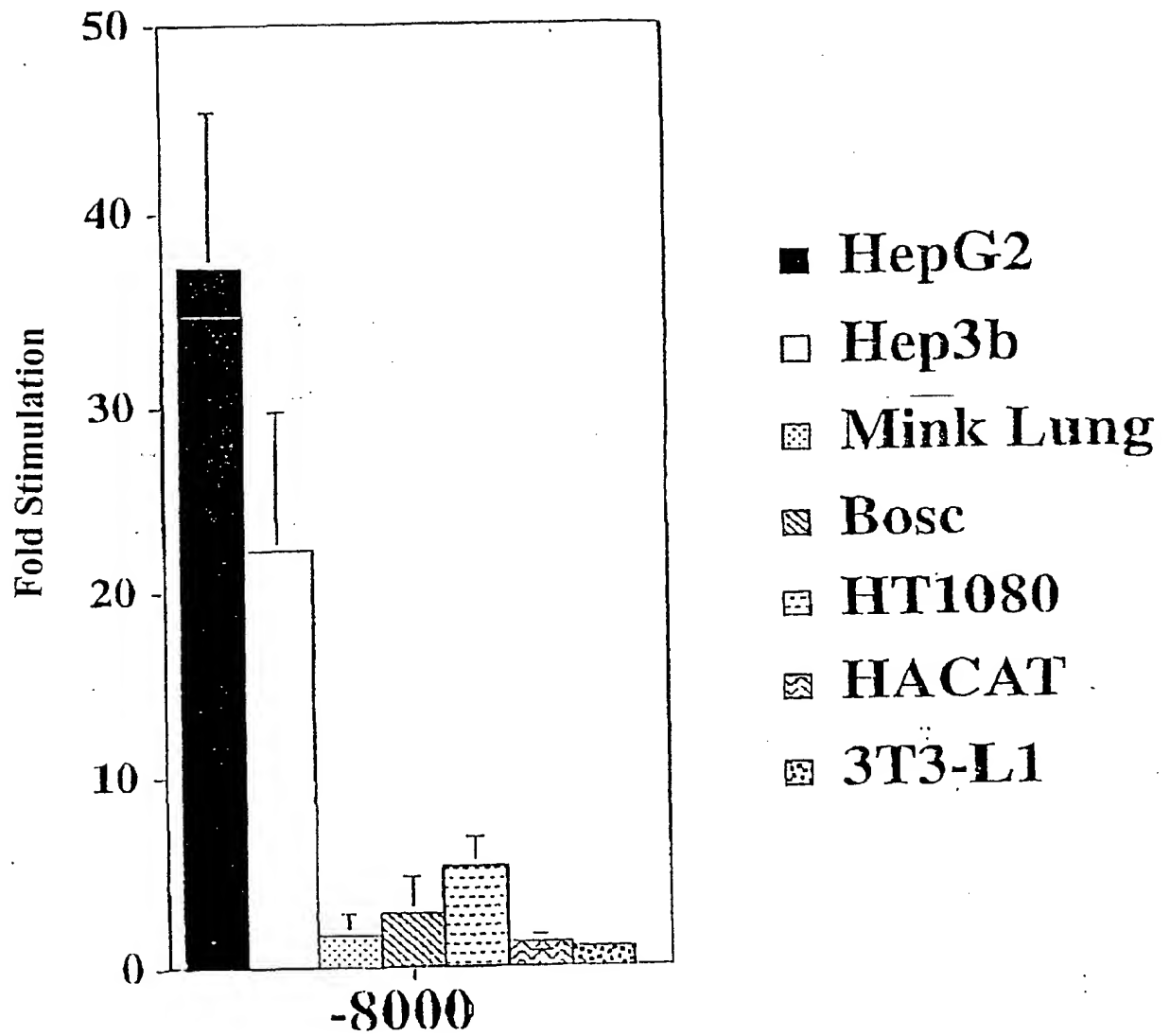


FIG. 100

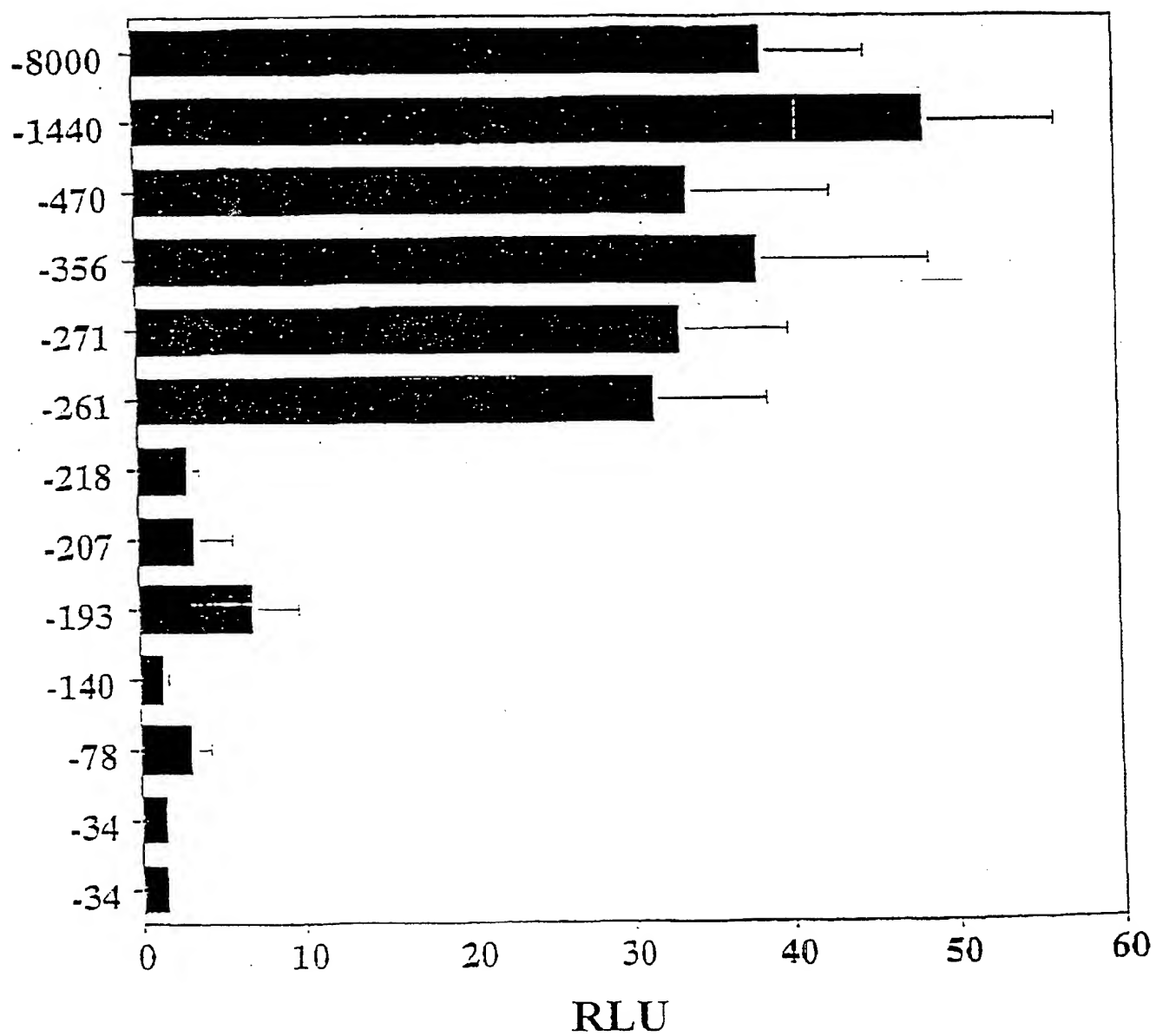


FIG. 101

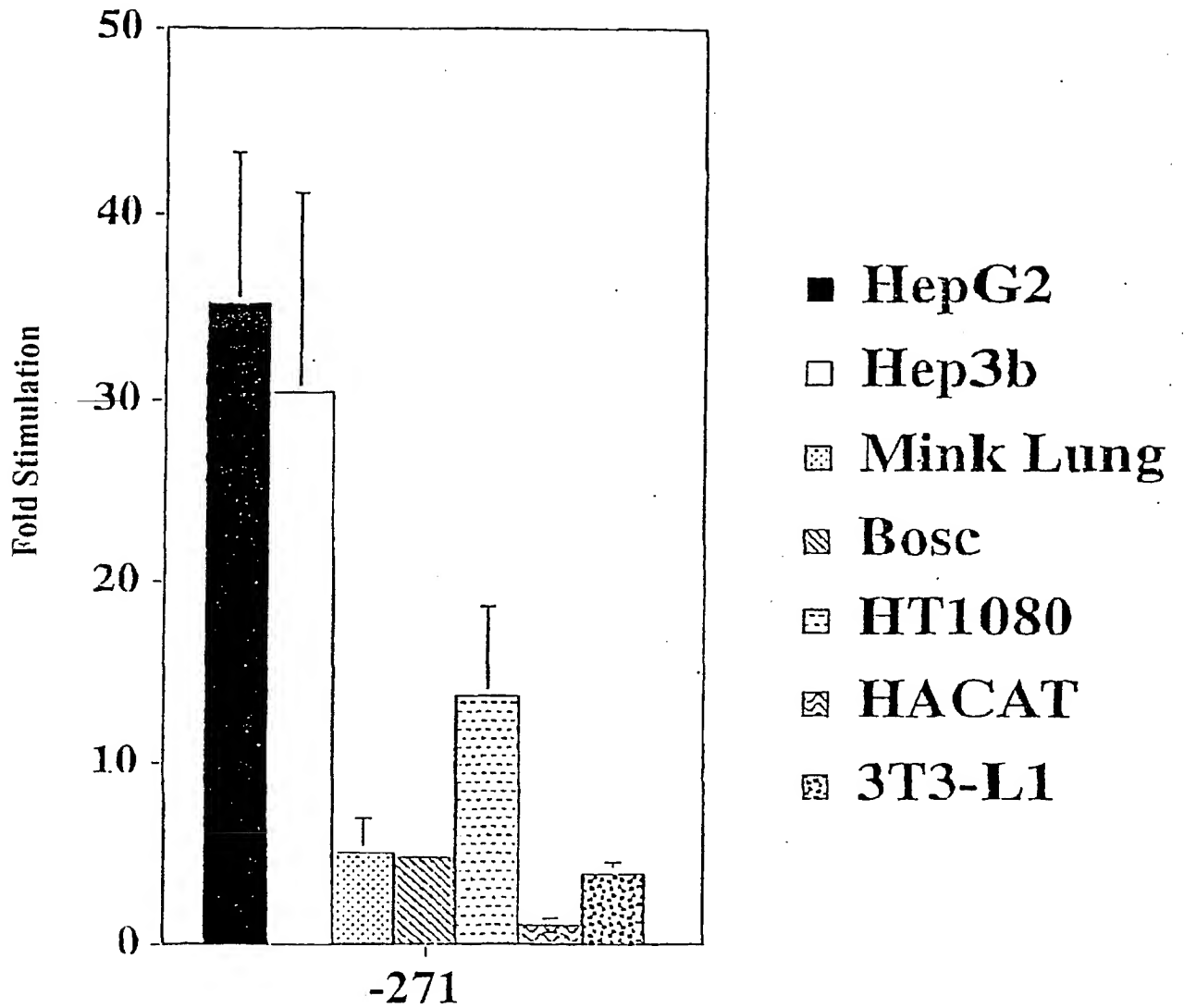


FIG. 102

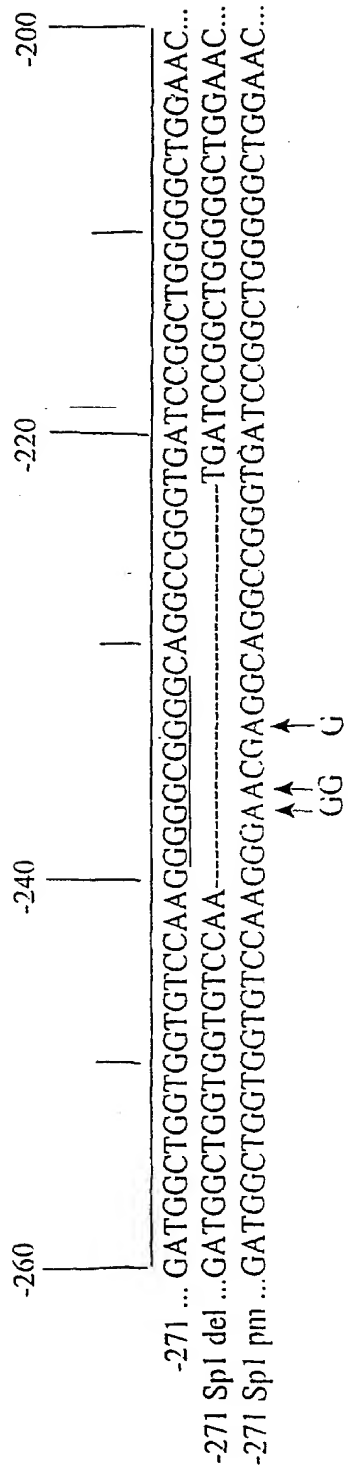


FIG. 103A

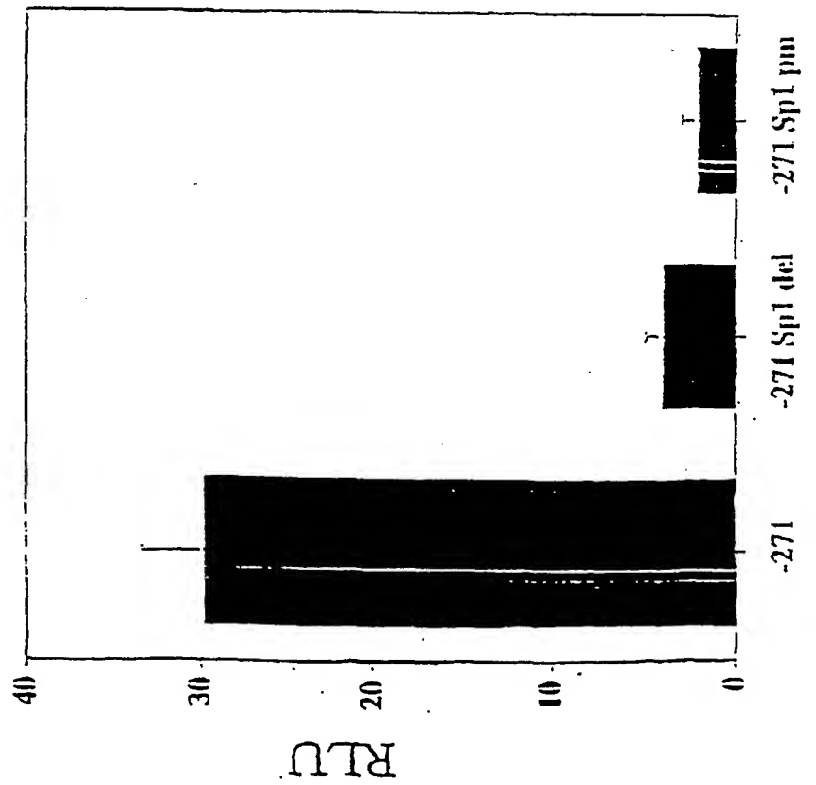


FIG. 103B



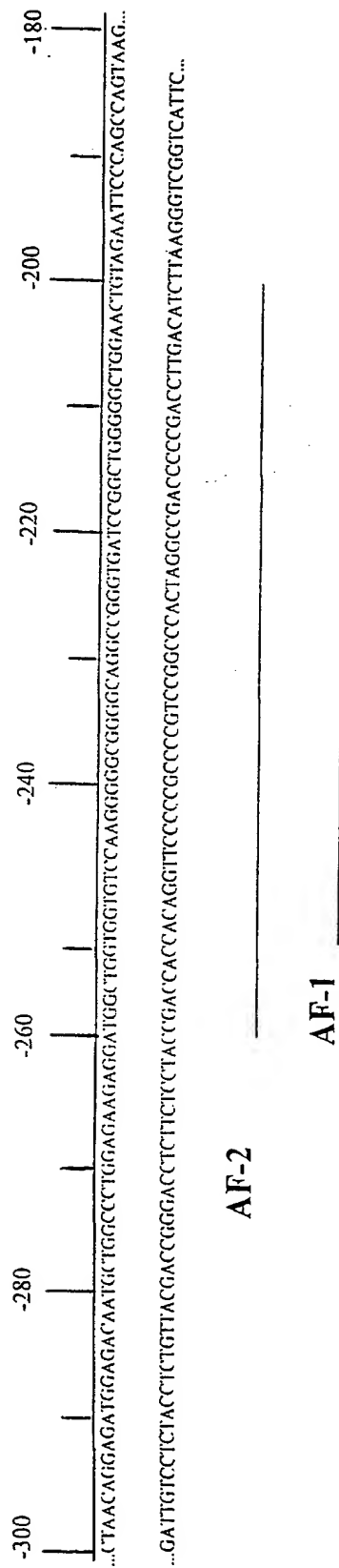


FIG. 104

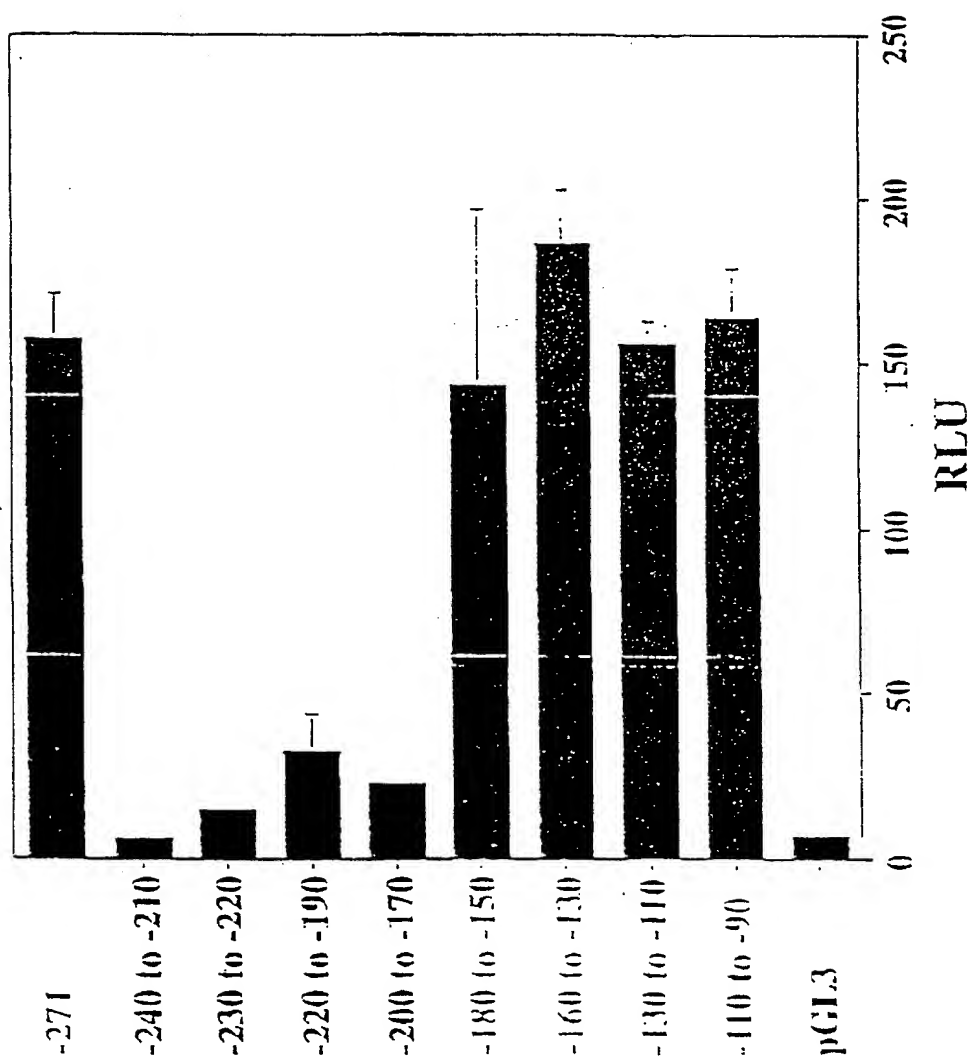


FIG. 105

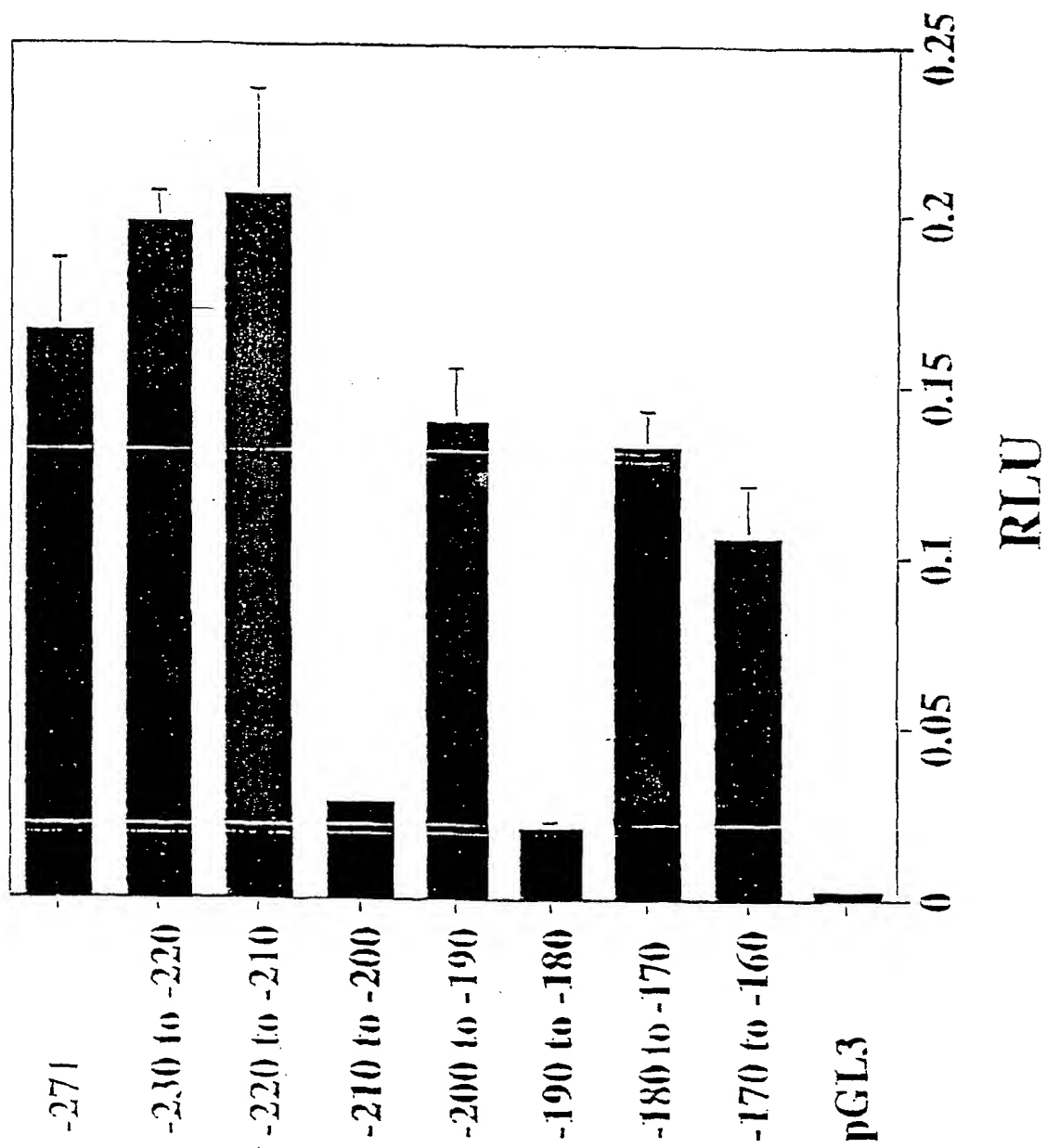


FIG. 106

GAGGATGGCTGGTGGTGTCCAAGGGGGCGGGGC

Sp1/GC box

GGCTGGA

Motif 1

GGCTGGA

Motif 2

CAGCCAGTAA

GA

ACTAAGTAACAAAAGG

ACAGAGTCCATGGGTCACATTCAGTTGCTGATAG

TACTTGGTCATATTTGGGAAGTGGGTAGACAGAT

TTCCTTAAAGGCAGGTAGTTAGGGCTTTGGAGCA

CTCATCAGAGCTAAGAGAGATTACACGCTCTCAT

CTACTTCAGAAAGAGCCAATGCCATG

FIG. 107

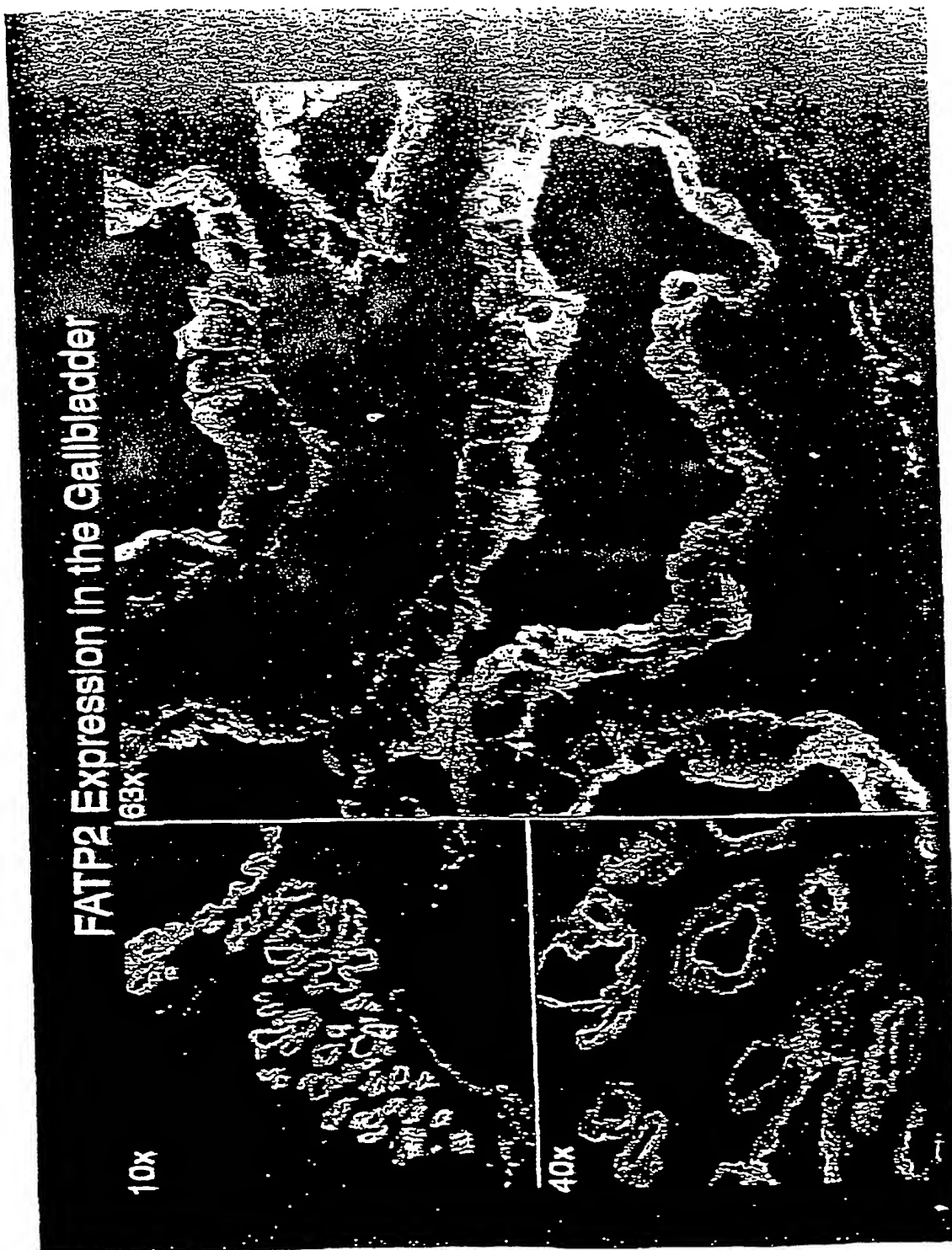


FIG. 108

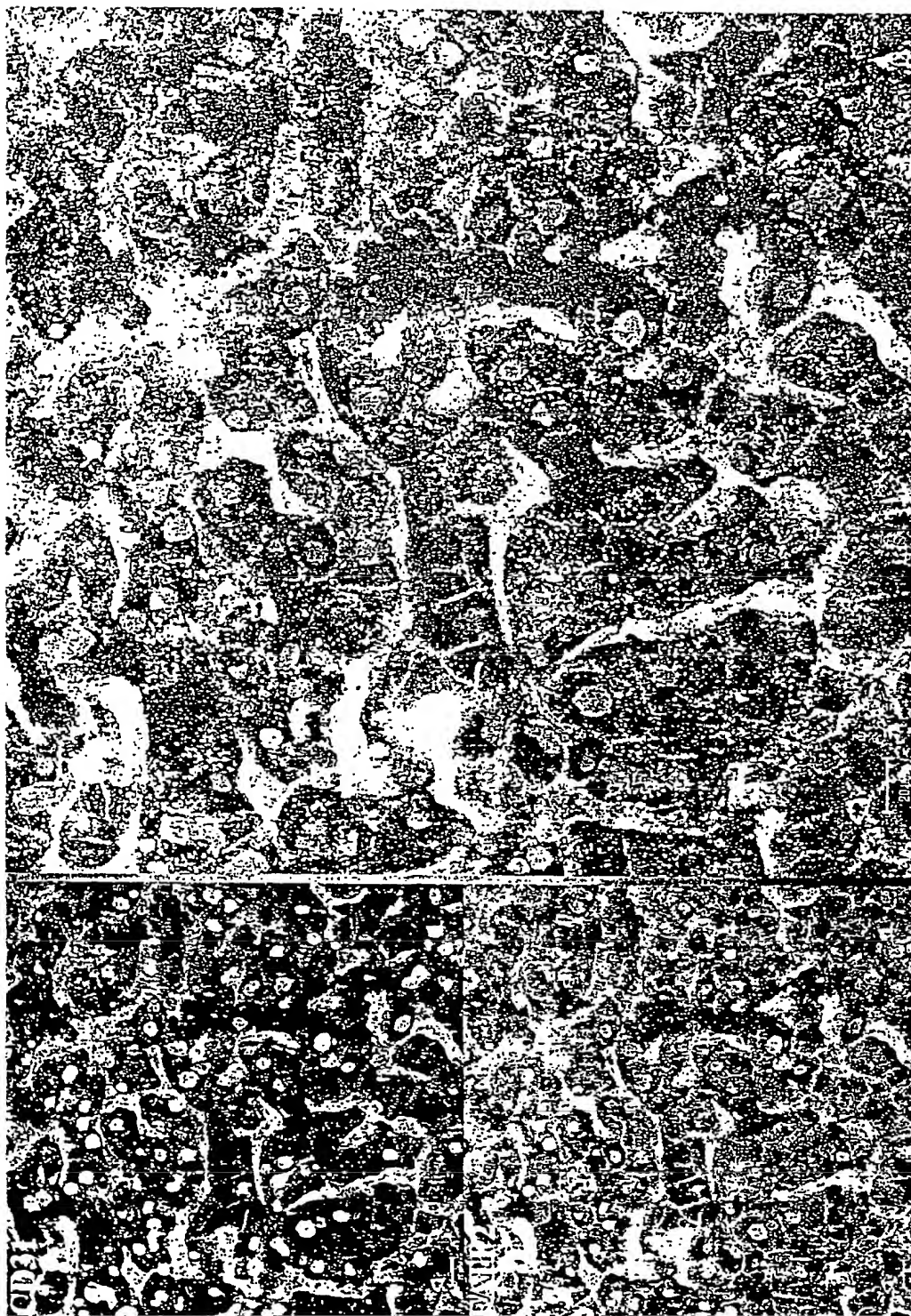


FIG. 109

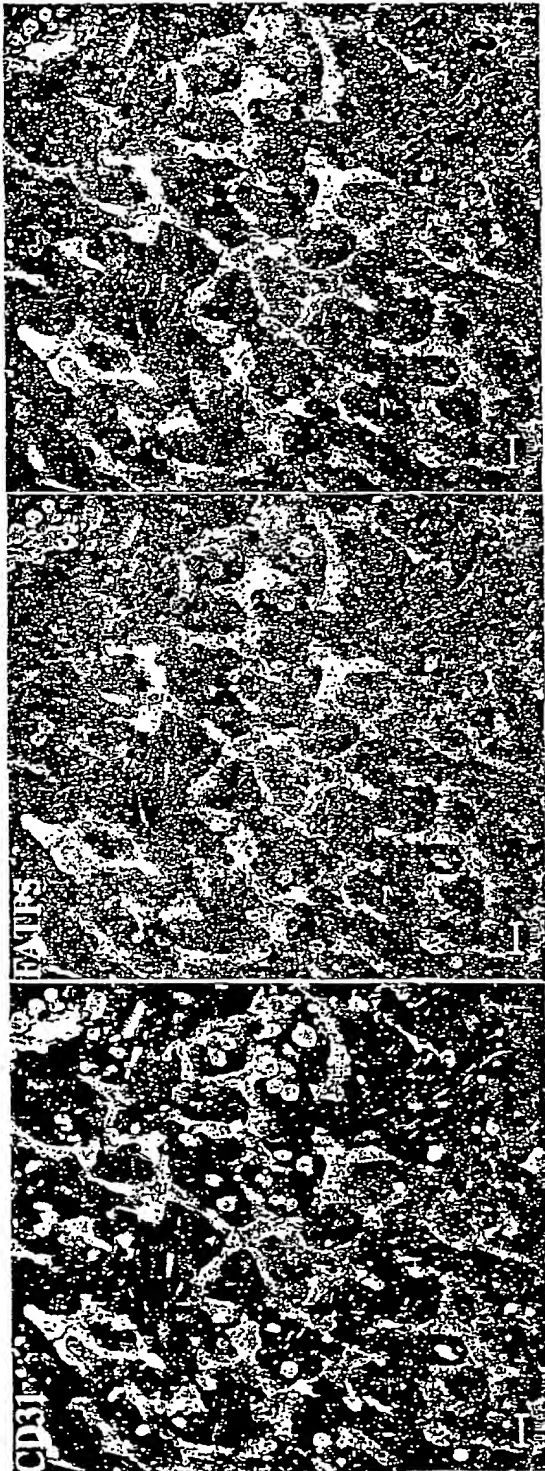


FIG. 110

**SUBSTITUTE SHEET (RULE 26)**

AA

FIG. 111



# FATP3 protein:

MGVCQRTAPWKEKSQLERAALGFRKGGGMFASGWNQTVPIEEAGSMAALLLLPLLLLLPLLLLLKLHLWPQLRWLPADL  
 AFAVRALCCKRALRARALAAAAADPEGEGCSLAWRLAELAQRAAHTFLIHGSRFFSYSEAEERESNRAARAFRLALGW  
 DWGPDGDSGEGSAGEGERAAPGAGDAAAGSGAEFAGDGAARGGGAAPLSPGATVALILPAGPEFLWLWFGGLAKAGLR  
 TAFVPTALRRGPLLHCLRSCGARALVLAPEFLESLEPDLPALRAMGLHLWAAGPGTHPAGISDLLAEVSAEVDGPVPGYL  
 SSPQSI TDTCLYIFTSGTTGLPKAARI SHLKILQCQGYQLCGVHQEDVIYLA LPLYHMSGSLLGIVGCMGIGATVVLKS  
 KFSAGQFWECCQQRVTVFYI GELCRYLVNQPPSKAERGHKVR LAVGSLRPDTWERFVRRFGPLQVLETYGLTEGNVA  
 TINYTGQRGAVGRASWLYKHIFPFSLIRYDVTTGEPIRD PQCHCMATSPGEPGLLIVAPVSQQSPFLGYAGGP ELAQGKLL  
 KDVFRPGDVFFNTGDL LVCDDQGF LRFHRTGDTFRWKGENVATTEVAEVEFALDFLQEVNVYGVTVPGHEGRAGMAALV  
 LRPPHALDLMQLYTHVSENLP PYARPRFLRLQESLATTETFKQQKVRMANEGFDPSTLS DPLYVLDDQAVGAYLPLTTARY  
 SALLAGNLRI

FIG. 112

BODIPY-FA uptake in cos cells

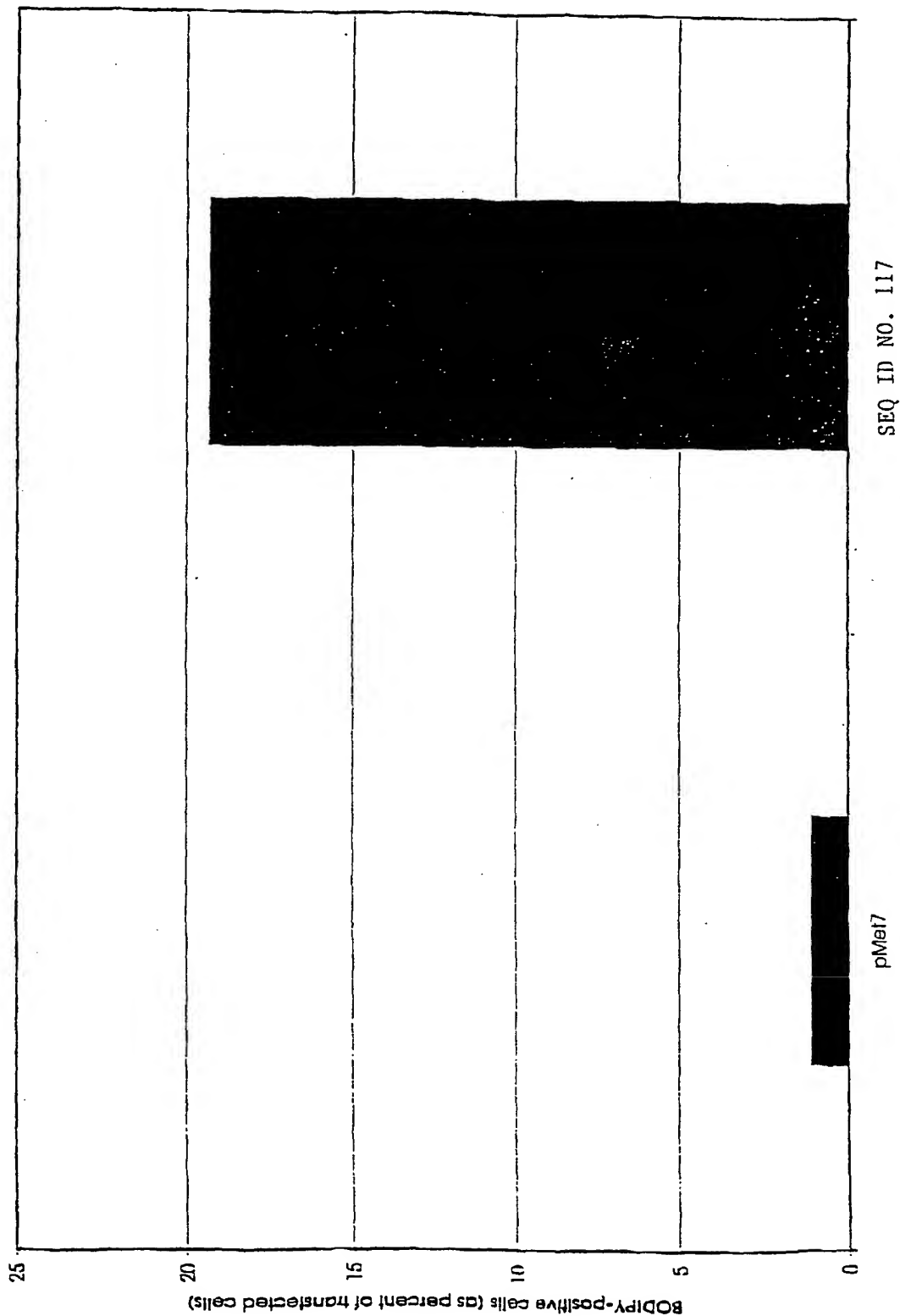


FIG. 113

Fig. 3

Peptide Sequence of hsFATP fragments for fatty acid binding experiments:

SP1:  
RVFIKTIIRDIFGGLVLLKVKAKVRQCLQERRTVPIFASTVRRHPDKTALIFEGTDTHWTFRQLDEYSSSVANFLQAR  
G  
LASGDVA AIFMENRNEFVGLWLGMAKLGVEAAALNTNLRDALLHCLTTSRARALVFGSEMASAICEVHASLDPSLS  
LFC  
SGSWEPGAVPPSTEHLDPLLKDA PKHLPSCPDKGFTD **FIG. 114A**

SP2:  
RVFIKTIIRDIFGGLVLLKVKAKVRQCLQERRTVPIFASTVRRHPDKTALIFEGTDTHWTFRQLDEYSSSVANFLQAR  
G  
LASGDVA AIFMENRNEFVGLWLGMAKLGVEAAALNTNLRDALLHCLTTSRARALVFGSEMASAICEVHASLDPSLS  
LFC  
SGSWEPGAVPPSTEHLDPLLKDA PKHLPSCPDKGFTDKLFYIYTSGTTGLPKAAIVVHSRYRMAALVYYGFRMRPN  
DIV  
YDCLPLYH **FIG. 114B**

SP3:  
GDVAAIFMENRNEFVGLWLGMAKLGVEAAALNTNLRDALLHCLTTSRARALVFGSEMASAICEVHASLDPSLSLFC  
SGS  
WEPGAVPPSTEHLDPLLKDA PKHLPSCPDKGFTDKLFYIYTSGTTGLPKAAIVVHSRYRMAALVYYGFRMRPN  
YDC  
LPLYH **FIG. 114C**

SP5:  
RLVRVNEDTMELIRGPDGVCPCQPGEPQLVGRIIQKDPLRRFDGYLNQGANNNKIAKDVFKKGDQAYLTGDVLM  
DEL  
GYLYFRDRTGDTFRWKGENVSTTEVEGTLRLLDMADVAVYGVVEVPGTEGRAGMAAVASPTGNCDLERFAQVLEK  
ELPLY  
ARPIFLRLPELHKGTGTYKFQKTELKKEGFDPAIVKDPFLFYLDQKGRYVPLDQEAYSRIQAGEEKL. **FIG. 114D**

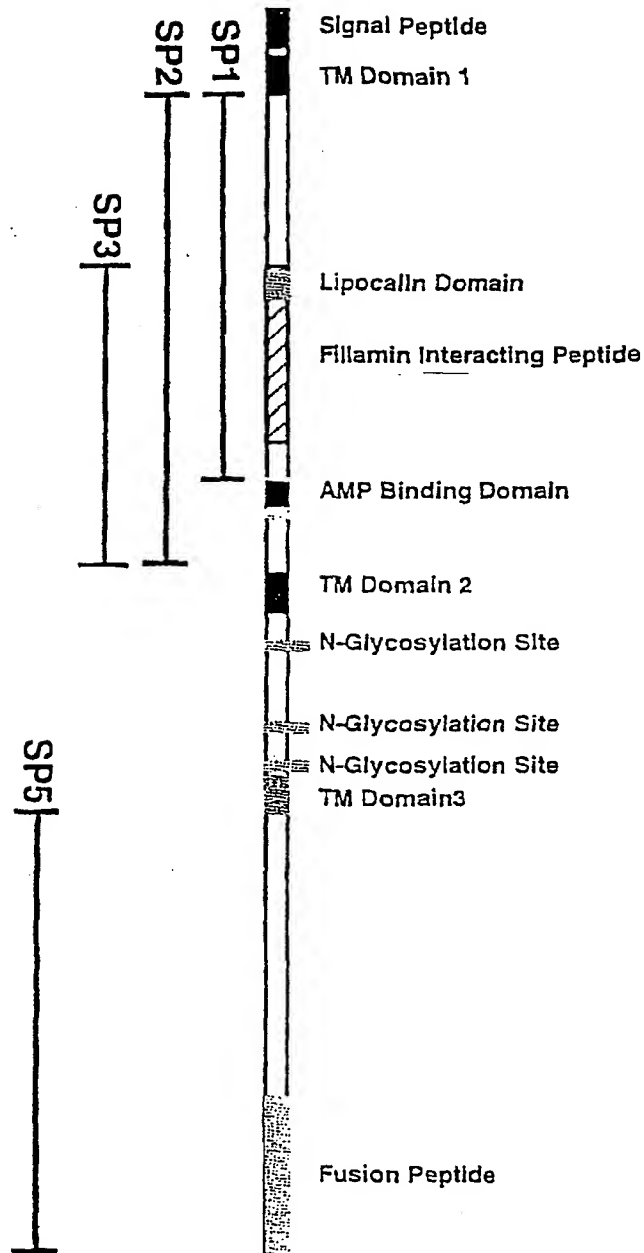


FIG. 115

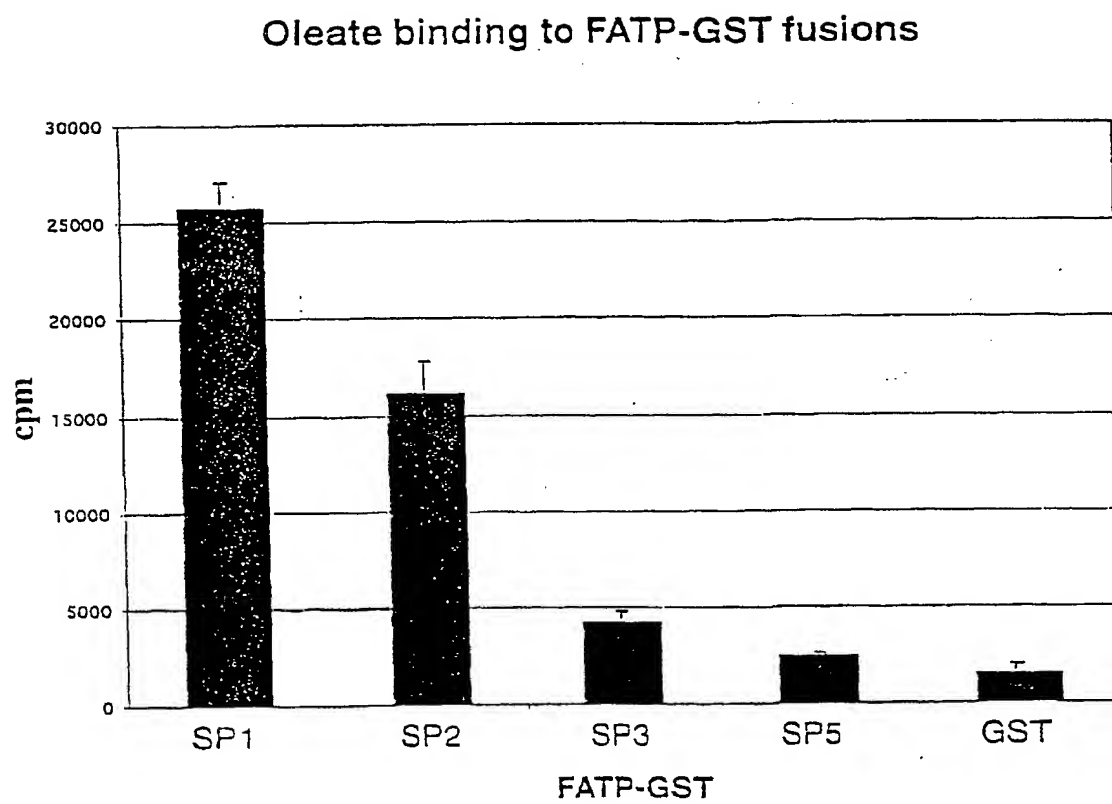


FIG. 116

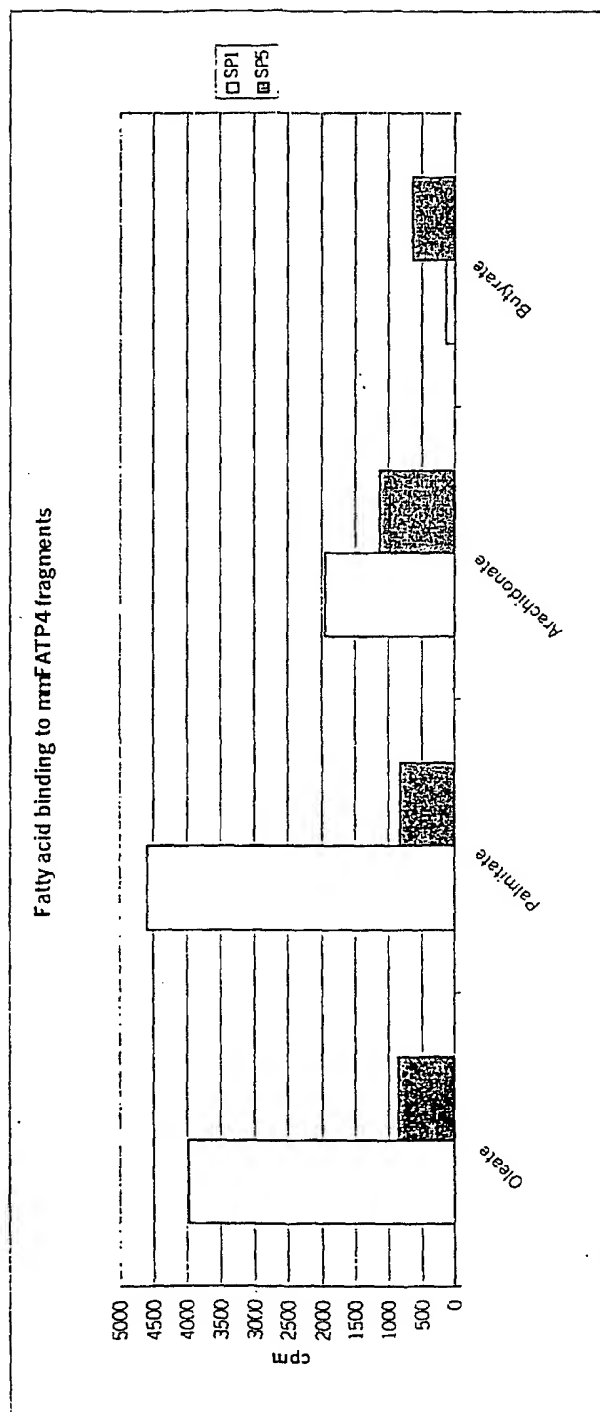


FIG. 117

**FIG. 118A**

hsFATP6..	56	V	L	D	K	F	L	S	H	A	K	R	Q	R	K	P	E	L	I	Y	E	G	D	-	-	-	I	Y	I	Y	Q	D	V	85		
hsFATP2..	55	I	L	R	A	F	L	E	K	A	R	Q	I	H	K	P	F	L	I	T	L	T	Y	A	Q	V	-	-	-	-	-	-	-	84		
hsFATP3..	114	L	A	W	R	L	A	E	L	A	Q	Q	H	A	H	T	F	L	I	H	G	S	R	-	-	-	-	-	-	-	-	-	-	143		
hsFATP5..	116	F	V	D	A	F	E	R	R	A	R	A	Q	P	P	F	A	L	L	V	W	T	G	P	G	A	G	S	V	T	F	G	E	L	148	
hsFATP4..	76	V	P	I	L	F	A	S	T	V	R	R	H	P	D	K	T	A	L	I	F	E	G	T	D	-	T	H	W	T	F	R	Q	L	107	
hsFATP1..	78	I	P	R	I	F	Q	A	V	I	V	Q	R	Q	R	E	R	L	A	L	V	D	A	G	T	G	-	E	C	W	I	F	A	Q	L	109
hsFATP6..	86	D	K	R	S	S	R	V	A	H	V	E	L	N	H	S	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	101	
hsFATP2..	85	D	R	R	S	N	Q	V	A	R	A	L	H	D	H	L	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100	
hsFATP3..	144	E	R	E	S	N	R	A	A	R	A	F	L	H	A	L	G	W	D	W	G	P	D	G	G	D	S	G	E	G	S	A	G	E	176	
hsFATP5..	149	D	A	H	A	C	Q	A	A	W	A	L	K	A	E	L	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	164	
hsFATP4..	108	D	E	Y	S	S	S	V	A	N	-	F	L	Q	A	R	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	122	
hsFATP1..	110	D	A	Y	S	N	A	V	A	N	-	L	F	H	Q	L	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	124	
hsFATP6..	102	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	101		
hsFATP2..	101	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100		
hsFATP3..	177	G	E	R	A	P	G	A	G	D	A	A	G	S	G	A	E	F	A	G	G	D	G	A	A	R	G	G	G	A	A	A	A	A	209	
hsFATP5..	165	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	167	
hsFATP4..	123	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	122	
hsFATP1..	125	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	124	

FIG. 118B



## LIPOCALIN DOMAIN

hsFATP6..	102	-	L	K	K	G	D	I	V	A	L	L	M	S	N	E	P	D	F	-	V	H	V	W	E	G	L	A	K	L	G	G	V	131	
hsFATP2..	101	-	L	R	Q	G	D	C	V	A	L	L	M	G	N	E	P	A	-	V	W	L	W	L	G	L	V	K	L	G	C	A	130		
hsFATP3..	210	P	L	S	P	G	A	T	V	A	L	L	L	P	A	G	P	E	F	-	L	W	L	W	F	G	L	A	K	A	G	L	R	240	
hsFATP5..	168	S	L	C	A	G	E	P	T	A	L	L	V	L	A	S	Q	A	V	P	A	L	C	M	W	L	G	L	A	K	L	G	C	P	200
hsFATP4..	123	-	L	A	S	G	D	V	A	A	I	F	M	E	N	R	N	E	F	-	V	G	L	W	L	G	M	A	K	L	G	V	E	152	
hsFATP1..	125	-	F	A	P	G	D	V	V	A	I	F	L	E	G	R	P	E	F	-	V	G	L	W	L	G	L	A	K	A	G	M	E	154	

hsFATP6..	132	V	A	F	L	N	T	N	I	R	S	N	S	L	L	N	C	I	R	A	C	G	P	R	A	L	V	V	G	A	D	L	L	G	164
hsFATP2..	131	M	A	C	L	N	Y	N	I	R	A	K	S	L	L	H	C	F	Q	C	C	G	A	K	V	L	L	V	S	P	E	L	Q	A	163
hsFATP3..	241	T	A	F	V	P	T	A	L	R	G	P	L	L	H	C	L	R	S	C	G	A	R	A	L	V	L	A	P	E	F	L	E	273	
hsFATP5..	201	T	A	W	I	N	P	H	G	R	G	M	P	L	A	H	S	V	L	S	S	G	A	R	V	L	V	V	D	P	D	L	R	E	233
hsFATP4..	153	A	A	L	I	N	T	N	L	R	H	D	A	L	L	H	C	L	T	T	S	R	A	R	A	L	V	F	G	S	E	M	A	S	185
hsFATP1..	155	A	A	L	L	N	V	N	L	R	R	E	P	L	A	F	C	L	G	T	S	G	A	K	A	L	I	F	G	G	E	M	V	A	187

hsFATP6..	165	T	V	E	E	I	L	P	S	L	S	-	E	N	I	S	V	W	G	M	K	-	D	S	V	P	Q	G	V	I	S	L	K	E	195
hsFATP2..	164	A	V	E	E	I	L	P	S	L	N	K	D	D	V	S	I	Y	Y	V	S	R	T	S	N	T	D	G	I	D	S	H	L	D	196
hsFATP3..	274	S	L	E	P	D	L	P	A	L	R	A	M	G	L	H	L	W	A	A	G	P	Q	T	H	P	A	G	I	S	D	L	L	A	306
hsFATP5..	234	S	L	E	E	I	L	P	K	L	Q	A	E	N	I	R	C	F	Y	L	S	H	T	S	P	T	P	G	V	G	A	L	G	A	266
hsFATP4..	186	A	I	C	E	V	H	A	S	L	D	-	P	S	L	S	L	F	C	S	G	-	S	W	E	P	G	A	V	P	P	S	T	E	216
hsFATP1..	188	A	V	A	E	V	S	G	H	L	G	-	K	S	L	I	K	F	C	S	G	-	D	L	G	P	E	G	I	L	P	D	T	H	218

## AMP-BINDING DOMAIN

hsFATP6..	196	K	L	S	T	S	P	D	E	P	V	P	R	S	H	H	V	V	S	L	-	L	K	S	T	C	L	Y	F	M	S	G	226	
hsFATP2..	197	K	V	D	E	V	S	T	E	P	I	R	E	S	W	R	S	E	V	I	-	-	F	S	T	P	A	L	Y	V	M	S	G	227
hsFATP3..	307	E	V	S	A	E	V	D	G	P	V	P	G	Y	L	S	S	P	Q	S	-	-	I	T	D	T	O	L	Y	F	M	S	G	337
hsFATP5..	267	A	L	D	A	A	P	S	H	P	V	P	A	D	L	R	A	G	I	T	-	-	W	R	S	P	A	L	Y	F	M	S	G	297
hsFATP4..	217	H	L	D	P	L	L	K	D	A	P	K	H	-	L	P	S	D	P	D	K	G	F	T	D	K	L	F	Y	F	M	S	G	248
hsFATP1..	219	L	L	D	P	L	L	K	E	A	S	T	A	P	L	A	Q	I	P	S	K	G	M	D	D	F	L	F	Y	F	M	S	G	251

FIG. 118C

AMP-BINDING DOMAIN  
CONTINUED

hsFATP6..	227	T	T	G	T	P	K	A	A	V	I	S	G	L	Q	V	L	R	G	S	A	V	L	W	A	E	G	-	C	T	A	H	D	I	258
hsFATP2..	228	T	T	G	T	P	K	A	A	M	T	H	Q	R	I	W	Y	G	T	G	L	T	F	V	S	G	I	-	L	K	A	D	D	V	259
hsFATP3..	338	T	T	G	T	P	K	A	A	R	I	S	H	L	K	I	L	Q	C	Q	G	F	Y	Q	L	C	G	-	V	H	Q	E	D	V	369
hsFATP5..	298	T	T	G	T	P	K	P	A	I	T	H	E	R	V	L	Q	M	S	K	M	L	S	L	S	G	-	A	T	A	D	D	V	329	
hsFATP4..	249	T	T	G	T	P	K	A	A	I	V	V	H	S	R	Y	Y	R	M	A	A	L	V	Y	Y	G	F	R	M	R	P	N	D	I	281
hsFATP1..	252	T	T	G	T	P	K	A	A	I	V	V	H	S	R	Y	Y	R	M	A	A	E	F	G	H	H	A	Y	R	M	Q	A	D	V	284
hsFATP6..	259	V	Y	I	T	L	P	L	Y	H	S	S	A	I	L	G	I	S	G	C	V	E	L	G	A	T	C	V	L	K	K	K	E	291	
hsFATP2..	260	I	Y	I	T	L	P	F	Y	H	S	A	A	L	I	G	I	H	G	C	I	V	A	G	A	T	L	A	L	R	T	K	F	292	
hsFATP3..	370	I	Y	L	A	L	P	L	Y	H	M	S	G	S	L	L	G	I	V	G	C	M	G	I	G	A	T	V	V	L	K	S	K	F	402
hsFATP5..	330	V	Y	T	V	L	P	L	Y	H	V	M	G	L	V	V	G	I	L	G	C	L	D	L	G	A	T	C	V	L	A	P	K	F	362
hsFATP4..	282	V	Y	D	C	L	P	L	Y	H	S	A	G	N	I	V	G	I	G	Q	C	L	L	H	G	M	T	V	V	I	R	K	K	F	314
hsFATP1..	285	L	Y	D	C	L	P	L	Y	H	S	A	G	N	I	L	G	V	G	Q	C	L	I	Y	G	L	T	V	V	L	R	K	K	E	317
hsFATP6..	292	S	A	S	Q	F	W	S	D	C	K	K	Y	D	V	T	V	F	Q	Y	I	G	E	L	C	R	Y	L	C	K	Q	S	K	R	324
hsFATP2..	293	S	A	S	Q	F	W	D	D	C	R	K	Y	N	V	T	V	I	Q	Y	I	G	E	L	L	R	Y	L	C	N	S	H	Q	K	325
hsFATP3..	403	S	A	G	Q	F	W	E	D	C	Q	Q	H	R	V	T	V	F	Q	Y	I	G	E	L	C	R	Y	L	V	N	I	P	P	S	435
hsFATP5..	363	S	T	S	C	F	W	D	D	C	R	Q	H	G	V	T	V	I	L	Y	V	G	E	L	L	R	Y	L	C	N	I	P	Q	Q	395
hsFATP4..	315	S	A	S	R	F	W	D	D	C	I	K	Y	N	C	T	I	V	Q	Y	I	G	E	L	C	R	Y	L	L	N	I	P	P	R	347
hsFATP1..	318	S	A	S	R	F	W	D	D	C	I	K	Y	N	C	T	V	V	Q	Y	I	G	E	L	C	R	Y	L	L	K	Q	P	V	R	350

FIG. 118D

hsFATP6..	325	EGE	K	D	H	K	V	R	L	A	I	G	N	G	I	R	S	D	V	W	R	E	E	L	D	R	F	G	N	I	K	V	357			
hsFATP2..	326	P	N	D	R	D	H	K	V	R	L	A	I	G	N	G	L	R	G	D	V	W	R	Q	F	V	K	R	F	G	D	I	C	I	358	
hsFATP3..	436	K	A	E	R	G	H	K	V	R	L	A	V	G	S	I	G	L	R	P	D	T	W	E	R	F	V	R	R	F	G	P	L	Q	V	468
hsFATP5..	396	P	E	D	R	T	H	T	V	R	L	A	M	G	N	G	L	R	A	D	V	W	E	T	F	Q	Q	R	F	G	P	I	R	I	428	
hsFATP4..	348	E	A	E	N	Q	H	Q	V	R	M	A	L	G	N	G	L	R	Q	S	I	W	T	N	F	S	S	R	F	H	I	P	Q	V	380	
hsFATP1..	351	E	A	E	R	R	H	R	V	R	L	A	V	G	N	G	L	R	P	A	I	W	E	F	T	E	R	E	G	V	R	Q	I	383		
hsFATP6..	358	C	E	L	Y	A	A	T	E	S	I	S	F	M	N	Y	T	G	R	I	G	A	I	G	R	T	N	L	F	Y	K	L	390			
hsFATP2..	359	Y	E	F	Y	A	A	T	E	G	N	I	G	F	M	N	Y	A	R	K	V	G	A	V	G	R	V	N	Y	L	Q	K	K	I	391	
hsFATP3..	469	L	E	T	Y	G	L	T	E	G	N	V	A	T	I	N	Y	T	G	Q	R	G	A	V	G	R	A	S	W	L	Y	K	H	I	501	
hsFATP5..	429	W	E	V	Y	G	S	T	E	G	N	M	G	L	V	N	Y	V	G	R	Q	G	A	L	G	K	M	S	C	L	L	R	M	L	461	
hsFATP4..	381	A	E	F	Y	G	A	T	E	C	N	C	S	L	G	N	F	D	S	Q	V	G	A	C	G	F	N	S	R	I	L	S	F	V	413	
hsFATP1..	384	G	E	F	Y	G	A	T	E	C	N	C	S	I	A	N	M	D	G	K	V	G	S	C	G	F	N	S	R	I	L	P	H	V	416	
hsFATP6..	391	S	T	F	D	L	I	K	Y	D	F	Q	K	D	E	P	M	R	N	E	Q	G	W	C	I	H	V	K	K	G	E	P	G	L	423	
hsFATP2..	392	I	T	Y	D	L	I	K	Y	D	V	E	K	D	E	P	V	R	D	E	N	G	Y	C	V	I	R	V	P	K	G	E	V	G	L	424
hsFATP3..	502	H	P	F	S	L	I	R	Y	D	V	T	T	G	E	P	I	R	D	P	Q	G	H	C	M	A	T	S	P	G	E	P	G	L	534	
hsFATP5..	462	S	P	F	E	L	V	Q	F	D	M	E	A	A	E	P	V	R	D	N	Q	G	F	C	I	P	V	G	L	G	E	P	G	L	494	
hsFATP4..	414	Y	P	I	R	L	V	R	V	N	E	D	T	M	E	L	I	R	G	P	D	G	V	C	I	P	C	Q	P	G	E	P	G	Q	446	
hsFATP1..	417	Y	P	I	R	L	V	K	V	N	E	D	T	M	E	L	L	R	D	A	Q	G	L	C	I	P	C	Q	A	G	E	P	G	L	449	
hsFATP6..	424	L	I	S	R	V	N	A	K	N	P	F	F	E	G	Y	A	G	-	-	P	Y	K	H	T	K	D	K	L	C	D	V	F	K	454	
hsFATP2..	425	L	V	C	K	I	T	Q	L	T	P	F	N	G	Y	A	G	-	-	A	K	A	Q	T	E	K	K	K	I	R	D	V	F	K	455	
hsFATP3..	535	L	V	A	P	V	S	Q	Q	S	P	F	L	G	Y	A	G	-	-	G	P	E	L	A	Q	G	K	L	L	K	D	V	F	R	565	
hsFATP5..	495	L	L	T	K	V	S	Q	Q	P	F	V	G	Y	R	G	-	-	P	R	E	L	S	E	R	K	L	V	R	N	V	R	Q	525		
hsFATP4..	447	L	V	G	R	I	I	Q	K	D	P	L	R	R	E	D	G	Y	L	N	Q	G	A	N	N	K	K	I	A	K	D	V	F	K	479	
hsFATP1..	450	L	V	G	Q	I	N	Q	Q	D	P	L	R	R	E	D	G	Y	V	S	E	S	A	T	S	K	K	L	A	H	S	V	F	S	482	

FIG. 118E

hsFATP6	455	K	G	D	V	Y	L	N	T	G	D	L	L	V	Q	D	N	E	L	Y	E	W	D	R	I	G	D	T	E	R	W	K	487		
hsFATP2.	456	K	G	D	L	Y	F	N	S	G	D	L	L	M	V	D	H	E	N	F	I	Y	H	D	R	V	G	D	T	F	R	W	K	488	
hsFATP3.	566	P	G	D	V	F	F	N	T	G	D	L	L	V	C	D	Q	G	F	L	R	F	H	D	R	T	G	D	T	F	R	W	K	598	
hsFATP5.	526	S	G	D	V	Y	Y	N	T	G	D	V	L	A	M	D	R	E	G	F	L	Y	F	R	D	R	L	G	D	T	F	R	W	K	558
hsFATP4.	480	K	G	D	Q	A	Y	L	T	G	D	V	L	V	M	D	E	L	G	Y	L	Y	F	R	D	R	T	G	D	T	F	R	W	K	512
hsFATP1.	483	K	G	D	S	A	Y	L	S	G	D	V	L	V	M	D	E	L	G	Y	M	Y	E	R	D	R	S	G	D	T	E	R	W	R	515
hsFATP6.	488	G	E	N	V	A	T	T	E	V	A	D	V	I	G	M	L	D	F	I	Q	E	A	N	V	Y	G	V	A	I	S	G	Y	E	520
hsFATP2.	489	G	E	N	V	A	T	T	E	V	A	D	T	V	G	L	V	D	F	V	Q	E	V	N	V	Y	G	V	H	V	P	D	H	E	521
hsFATP3.	599	G	E	N	V	A	T	T	E	V	A	E	V	F	E	A	L	D	F	L	Q	E	V	N	V	Y	G	V	T	V	P	G	H	E	631
hsFATP5.	559	G	E	N	V	S	T	H	E	V	E	G	V	L	S	Q	V	D	F	L	Q	Q	V	N	V	Y	G	V	C	V	P	D	C	E	591
hsFATP4.	513	G	E	N	V	S	T	T	E	V	E	G	T	L	S	R	L	D	M	A	D	V	A	V	Y	G	V	E	V	P	G	T	E	545	
hsFATP1.	516	G	E	N	V	S	T	T	E	V	E	G	V	L	S	R	L	L	D	Q	T	D	V	A	V	Y	G	V	A	V	P	G	V	E	548
hsFATP6.	521	G	R	A	G	M	A	S	I	L	L	K	P	N	I	S	L	D	L	E	K	V	Y	E	Q	V	V	T	E	L	P	A	Y	A	553
hsFATP2.	522	G	R	I	G	M	A	S	I	K	M	K	E	N	H	E	F	D	G	K	K	L	R	Q	H	I	A	D	Y	L	P	S	Y	A	554
hsFATP3.	632	G	R	A	G	M	A	A	L	V	L	R	P	P	H	A	L	D	L	M	Q	L	Y	T	H	V	S	E	N	L	P	P	Y	A	664
hsFATP5.	592	G	K	V	G	M	A	A	V	Q	L	A	P	Q	T	F	D	G	E	K	L	M	Q	H	V	R	A	W	L	P	A	Y	A	624	
hsFATP4.	546	G	R	A	G	M	A	A	V	A	S	-	P	T	G	N	C	D	L	E	R	F	A	Q	V	L	E	K	E	L	P	L	Y	A	577
hsFATP1.	549	G	K	A	G	M	A	A	V	A	D	-	P	H	S	L	L	D	P	N	A	I	Y	Q	E	L	Q	K	M	L	A	P	Y	A	580

FIG. 118F

hsFATP6..	554	C	P	R	F	L	R	I	Q	E	K	M	E	A	T	G	I	F	E	K	L	L	K	H	Q	L	V	E	D	G	F	N	P	L	K	586
hsFATP2..	555	R	P	R	F	L	R	I	Q	D	T	I	E	I	T	G	T	F	K	H	R	K	M	T	L	V	E	E	G	F	N	P	A	V	587	
hsFATP3..	665	R	P	R	F	L	R	I	Q	E	S	L	A	T	T	E	T	F	K	Q	K	V	R	M	A	N	E	G	F	D	P	S	T	697		
hsFATP5..	625	T	P	H	F	I	R	I	Q	D	A	M	E	V	T	S	T	F	K	L	M	K	T	R	L	V	R	E	G	F	N	V	G	I	657	
hsFATP4..	578	R	P	I	F	L	R	L	P	E	L	H	K	T	G	T	Y	K	F	Q	K	T	E	L	R	K	E	G	F	D	P	A	I	610		
hsFATP1..	581	R	P	I	F	L	R	L	P	Q	V	D	T	I	G	I	F	E	K	I	Q	K	T	R	L	Q	R	E	G	E	D	P	R	Q	613	
hsFATP6..	587	I	S	E	P	L	Y	F	E	M	D	N	L	K	K	S	Y	V	L	L	T	R	E	L	Y	D	Q	I	M	L	G	E	I	K	L	619
hsFATP2..	588	I	K	D	A	L	Y	F	L	D	D	T	A	K	M	Y	V	P	M	T	I	E	D	I	Y	N	A	I	S	A	K	T	L	K	L	620
hsFATP3..	698	L	S	D	P	L	Y	V	L	D	Q	A	V	G	A	Y	L	P	L	T	T	A	R	Y	S	A	L	L	A	G	N	L	R	I	730	
hsFATP5..	658	V	V	D	P	L	F	V	L	D	N	R	A	Q	S	F	R	P	L	T	A	E	M	Y	Q	A	V	C	E	G	T	W	R	L	690	
hsFATP4..	611	V	K	D	P	L	F	Y	L	D	A	Q	K	G	R	Y	V	P	L	D	Q	E	A	Y	S	R	I	Q	A	G	E	E	K	L	643	
hsFATP1..	614	T	S	D	R	L	F	F	L	D	L	K	Q	G	H	Y	L	P	L	N	E	A	V	Y	T	R	I	C	S	G	A	F	A	L	646	

FIG. 118G

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/25891

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/12 C07K14/705 C12N5/10 C12N15/62 G01N33/50  
G01N33/68 C12Q1/68 C07K16/28 A61K47/48  
//A61P3/04

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, BIOSIS

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 99 36537 A (TARTAGLIA LOUIS A ;WHITEHEAD BIOMEDICAL INST (US); GIMENO RUTH E () 22 July 1999 (1999-07-22) cited in the application	1-43, 45, 47, 49, 50, 52-66, 70-76, 79, 81-86, 88, 90, 92, 94, 96, 119, 120, 123-125
Y	the whole document	70-86, 88, 90, 92, 94, 96, 122-125
	---	---
	--/--	

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

### \* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

15 February 2001

Date of mailing of the international search report

10.05.01

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040. Tx. 31 651 epo nl.  
Fax: (+31-70) 340-3016

Authorized officer

ANDRES S.M.

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/25891

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	FLOWER D: "The lipocalin protein family: structure and function" BIOCHEMICAL JOURNAL, vol. 318, 1996, pages 1-14, XP002095126 ISSN: 0264-6021 the whole document	70-86, 88,90, 92,94, 96, 122-125
X	WO 99 46281 A (BAKER KEVIN P ;CHEN JIAN (US); GENENTECH INC (US); GURNEY AUSTIN () 16 September 1999 (1999-09-16)  abstract page 10, paragraph 15. page 275 -page 276; example 114 figure 39 claims	3,31, 33-42, 55,61, 62, 70-76, 81-85
A	HIRSCH ET AL: "A family of fatty acid transporters conserved from mycobacterium to man" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 95, July 1998 (1998-07), pages 8625-8629, XP002107853 ISSN: 0027-8424 cited in the application	
A	BANASZAK, L. ET AL.: "Lipid-binding proteins: A family of fatty acid and retinoid transport proteins." ADVANCES IN PROTEIN CHEMISTRY, vol. 45, 1994, pages 89-151, XP000982080 ISBN: 0-12-034245-6 cited in the application	
X	DATABASE EM_EST [Online] EMBL; ID: HSAA13688; Accession number : AA193416, 23 January 1997 (1997-01-23) HILLIER, L. ET AL.: "zr40b06.r1 Soares NhMPu S1 Homo sapiens cDNA clone IMAGE:665843 5'" XP002160453 abstract & HILLIER, L. ET AL.: "Generation and expression of 280,000 human expressed sequence tags" GENOME RESEARCH, vol. 6, 1996, pages 807-828, XP000914615	4-6,29, 30

-/--

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 00/25891

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE EM_EST [Online]  EMBL;  Accession number : AI041230,  1 July 1998 (1998-07-01)  "ov56b02.x1 Soares testis NHT Homo  sapiens cDNA clone IMAGE:1641291 3'"  XP002160454  abstract</p> <p>-----</p>	<p>4-6,29,  30</p>



# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US 00/25891

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☒ Claims Nos.: 44 46 48 51 87 89 91 93 95 97 121  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:  
see FURTHER INFORMATION sheet PCT/ISA/210
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:  
1-66, 119-121 (in totality) and 67-97, 122-125 (all partially)

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-66,119-121 (in totality) and 67-97,  
122-125 (all partially)

Human fatty acid transporter protein FATP3 characterised by SEQ IDs 116 and 117; fragments and variants thereof; fusion proteins, vectors or cells comprising them; antibodies against them; their use in methods for screening modulating agents, for diagnostic or therapy.

2. Claims: 67-97,122-125 (all partially)

As for subject 1, but concerning FATP1 characterised by SEQ IDs 46 and 47.

3. Claims: 67-97,101,122-125 (all partially) and 102-106,  
112-116 (totally)

As for subject 1, but concerning FATP2 characterised by SEQ IDs 48 and 49. Methods for directing an agent to liver cells or gall bladder using a moiety which binds to FATP2.

4. Claims: 67-97,122-125 (all partially)

As for subject 1, but concerning FATP4 characterised by SEQ IDs 52 and 53.

5. Claims: 67-97,101,122-125 (all partially) and 98-100,  
107-111,117-118 (totally)

As for subject 1, but concerning FATP5 characterised by SEQ IDs 54 and 55. A promoter sequence derived therefrom and a method for directing an agent to liver cells using a moiety which binds to FATP5.

6. Claims: 67-97,122-125 (all partially)

As for subject 1, but concerning FATP6 characterised by SEQ IDs 56 and 57.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 44 46 48 51 87 89 91 93 95 97 121

Claims 44, 46, 48, 51, 87, 89, 91, 93, 95, 97 and 121 relate to compounds defined solely by a desired property or a method of identification. As there is no structural feature characterising these compounds and allowing a meaningful search, these claims have not been searched.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 00/25891

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9936537 A	22-07-1999	AU 2310899 A	02-08-1999
		EP 1045904 A	25-10-2000
WO 9946281 A	16-09-1999	AU 3072199 A	27-09-1999
		AU 3075099 A	11-10-1999
		EP 1064382 A	03-01-2001
		WO 9947677 A	23-09-1999
		AU 1532499 A	15-06-1999
		EP 1032667 A	06-09-2000
		WO 9927098 A	03-06-1999
		AU 3757099 A	08-11-1999
		EP 1071773 A	31-01-2001
		WO 9954467 A	28-10-1999
		AU 1070399 A	10-05-1999
		EP 1025227 A	09-08-2000
		WO 9920756 A	29-04-1999